JAN 24 1921
UNIV. OF MICH.
JAN 24 1921

Journal of Educational Research

Official Organ of the National Association of Directors of Educational Research

B. R. BUCKINGHAM, *Editor*
Associate Editors
E. J. Ashbaugh — W. W. Charters — George D. Strayer
Leonard P. Ayres — J. A. Courtis — Lewis M. Terman
Walter S. Monroe

VOL. II — DECEMBER, 1920 — No. 5

CONTENTS

A Group Intelligence Examination without Prepared Blanks, *J. Crosby Chapman* 777
Suggestions for Procedure Following a Testing Program: I. Reclassification, *B. R. Buckingham* 787
What Should Teacher-Rating Schemes Seek to Measure? *Raymond A. Kent* 803
Grade Norms for the New York City Penmanship Scale, *Eugene A. Nifenecker* 808
A Comparative Study of the Intelligence of White and Colored Children, *H. A. Schweiger* and *Edith Winn* 838
Educational Measurement in the Field of History, *Chas. L. Harlan* 849
Editorials 854
Reviews and Abstracts 857
News Items and Communications 860
National Association of Directors of Educational Research 871

Entered as second-class matter January 15, 1920, at the Post Office at Menasha, Wisconsin, under the act of March 3, 1879

Journal of Educational Research

Copyright 1921, by
Public School Publishing Co.

Published for the Bureau of Educational Research, University of Illinois

Editor-in-Chief: B. R. Buckingham
Director of Bureau of Educational Research, University of Illinois

Associate Editors

E. J. ASHBAUGH, *Ex-Officio, Secretary of Research Association, University of Iowa, Iowa City*
LEONARD P. AYRES,
Vice-President, Cleveland Trust Company, Cleveland, Ohio.
W. W. CHARTERS,
Carnegie Inst. Tech., Pittsburgh
S. A. COURTIS,
City Schools, Detroit, Mich.
WALTER S. MONROE,
University of Illinois, Urbana, Ill.
GEORGE D. STRAYER,
Columbia University, New York City
LEWIS M. TERMAN,
Leland Stanford Junior University, Stanford University, Calif.

Contributing Editors

VIRGIL E. DICKSON,
Director, Bureau of Research and Guidance, Public Schools Oakland and Berkeley, California
CHARLES FORDYCE,
Dean Teachers' College, and Director of Bureau of Educational Measurements, University of Nebraska.
WILLIAM S. GRAY,
Dean College of Education, University of Chicago.
M. E. HAGGERTY,
Dean of the College of Education, University of Minnesota.
V. A. C. HENMON,
Director School of Education, University of Wisconsin.
ARTHUR W. KALLOM,
Assistant Director Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.
F. J. KELLY,
Dean, School of Education and Director Bureau of School Service, University of Kansas.
ROBT. H. LANE,
Assistant Superintendent of Schools, Los Angeles Calif.
WILLIAM A. McCALL,
Assistant Professor of Education, Teachers' College, Columbia University.
GEORGE W. MELCHER,
Director Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.
EUGENE A. NIFENECKER,
Assistant Director Bureau of Reference, Research, and Statistics, Public Schools, New York City.
JOSEPH P. O'HERN,
Assistant Superintendent of Schools and Director of Efficiency Bureau, Public Schools, Rochester New York.
LEON O. SMITH,
Assistant Superintendent (in charge of Educational Research) Public Schools, Omaha, Nebraska.
W. W. THEISEN,
Director Division of Reference and Research, Public Schools, Cleveland, Ohio.
CLIFFORD WOODY,
Professor of Education, University of Washington.
E. H. CAMERON,
Editor of "Reviews and Abstracts," University of Illinois, Urbana, Ill.

THE JOURNAL OF EDUCATIONAL RESEARCH is published at 450 Ahnaip St., Menasha, Wis. monthly except July and August by Geo. Banta, Official Printer to the Public School Publishing Co. The subscription price is $4.00; $4.30 in Canada or Mexico; and $4.50 in other foreign countries.

The editorial office is at Urbana, Illinois. Address all communications on educational matters, articles for the magazine etc. to B. R. Buckingham, University of Illinois, Urbana, Illinois.

The JOURNAL OF EDUCATIONAL RESEARCH is both needed by and needs the subscriptions of many more teachers and school officers.

The title page and index for Volume II are furnished with the December, 1920, number.

The Public School Publishing Company conducts the subscription work. Send all orders to them at Bloomington, Illinois.

Articles to Appear Soon

Development of the "Attainments in Reading" Chart by *J. P. O'Hern.*

Some Recent Results Obtained from the Otis Group Intelligence Tests by *Stephen S. Colvin.*

Provision in the High School Curriculum for Correcting Physical Defects by *Ernest McCleary Todd.*

Standardized Vocabulary and Sentence Test in French by *V. A. C. Henmon.*

Inequality of Educational Opportunity: A Study of the Schools in a Michigan Village by *Gilbert L. Brown.*

Journal of Educational Research

Editorial Office, Urbana, Illinois — Editor-in-chief, B. R. Buckingham, University of Illinois

VOLUME II — DECEMBER, 1920 — NUMBER 5

A GROUP INTELLIGENCE EXAMINATION WITHOUT PREPARED BLANKS

J. CROSBY CHAPMAN
Yale University

The extensive use of group intelligence tests bears witness to the values which can be derived from their application. All the tests in present use necessitate the purchase of prepared blanks for each pupil. Those of Otis, Thorndike, Haggerty, Thurstone, Dearborn, and Chapman and Welles may be cited as examples. The cost of administering these tests varies from approximately four cents to eighty cents per pupil. While those who are acquainted with the tests would readily assert that a few cents for each individual examined could hardly be spent to better advantage, yet taking conditions as they exist many school systems hesitate to use them owing to the financial outlay. For example in a situation which came to the author's attention where it was distinctly desirable to test sixteen hundred children, the authorities did not see their way clear to spend the necessary $80 to $160 for blanks which could only be used on a single occasion.

The question has often occurred to the author as to whether it was possible to construct an examination: (1) which would be reasonably reliable and valid; (2) which could be given to groups of thirty to fifty children; (3) which would involve merely paper and pencil; (4) which could be corrected by having the pupils exchange papers; and (5) which would not consume more than one hour for administering and correcting. This study briefly describes a tolerably accurate instrument[1] which has been devised to meet these conditions for upper grades and for high school. In view of the fact that the major portion of group testing concerns itself merely with the differences within the group and not with external comparison, the present form of examination may well be used in cases where for one reason or another it is not feasible

[1] Copies of this test, in any number, can be obtained, for twenty-five cents, from the author at the Department of Education, Yale University, New Haven, Connecticut.

to test in the more expensive way. The meticulously scientific reader may console himself with the thought that it is obviously better to use a slightly less flexible and exact instrument than no objective measure at all.

The first section presents a very short description of the construction of the test together with full directions for administering and scoring. In the second section there will be found statistical and other justification for the use of an examination of this type.

SECTION I

The test is made up of one hundred elements or items distributed as follows:

Information	25
Arithmetic	25
Opposites	25
Sentence Understanding	25

Other kinds of test, particularly directions and various forms of memory, were experimented with but they failed to justify themselves.

The elements of each of the tests were tried out on two groups: (1) the brightest ten children of grade VIII in a typical school containing 120 children in the grade; (2) the dullest ten among the same 120 children. The grouping was based on a combination of teacher estimates and school marks. Every test element that failed to be answered by some of the brightest group, that was answered by all the lower group, or answered by a greater or equal proportion of the duller than of the brighter group was eliminated. Each part of the test consisted finally of 25 elements, of various degrees of difficulty arranged at random. The time allowances were eventually fixed after the author and others had derived experience from giving the test to several hundred children.

General Directions

Preparation

Use paper ruled on both sides.

Instruct the pupils to fold the paper twice from top to bottom thereby forming three equal rectangles.

Instruct the pupils to number, line by line, from one to a hundred, passing down the whole length of one rectangle before going on to the next one and using both sides of the sheet. Have

the numbering done as near the edge as possible in order to leave the maximum amount of room for writing answers.

Explanation

1. Explain to the pupils the general nature of the test.

2. Tell them to do their best, and also inform them that the class is going to mark the papers at the close of the examination. Warn the pupils to keep their papers from the view of those around them.

3. In giving the test read each question as carefully as you can in a manner calculated to give the children the best possible chance.

4. If the pupils are writing the answer to the previous question do not attempt to make them stop writing as you dictate the next question. Go straight ahead with the examination. The pupil who falls behind and works slowly will handicap himself if he gets so much behind that he cannot give his undivided attention to the coming question.

5. Let the whole procedure be as automatic as possible. It is well to practice on a small group before trying large groups.

I. Information

Tell the pupils that 25 simple questions are to be asked, and that to each one of these questions they must write the answer, one word or figure as quickly as possible. Great care must be taken to see that the answer is written opposite the number which corresponds to the question. If they do not know the answer, **they must draw a dash** and wait for the next question. Dictate each question *once and once only*.

Give the pupils clearly to understand that the question will be given once, and that if they do not get it, they must not hold up their hands but draw a dash and wait for the next question. Tell them not to get discouraged if at any time they get muddled, but simply to wait for the next question.

After dictating a question let ten seconds elapse measured by counting ten to one's self at an even rate previously learnt by counting against a watch. (Note. Most examiners will tend to count too rapidly unless they check up their speed before they start by counting with a watch.) Then proceed to the next question. Always call the number of the question with the question itself.

INFORMATION TEST

1. In what month does Independence Day come?
2. What is the commonest fuel used in the automobile?
3. Who wrote "Evangeline"?
4. How many quarts are there in a peck?
5. What general surrendered at Appomattox?
6. In what country are the Pyramids?
7. From what animal does venison come?
8. What is the commonest make of automobile in the U. S.?
9. With what instrument is temperature usually measured?
10. How many amendments are there to the Constitution?
11. How many red stripes are there in the U. S. flag?
12. In what year will the next leap year come?
13. Into what tree does the acorn grow?
14. In what month was the armistice signed?
15. Who was the commanding admiral of the U. S. Navy during this last war?
16. What is the plural of tooth?
17. In the Bible story, whom did David kill by throwing a stone?
18. What is the feminine of shepherd?
19. For what article of consumption is Postum a substitute?
20. What is the normal temperature of the body?
21. What famous product is made by Wrigley?
22. How do you write 112 in Roman numerals?
23. Who is supposed to have discovered wireless telegraphy?
24. What is the technical name applied to a foreigner residing in the U. S., who is not naturalized?
25. What do the letters "A.B.," written after a man's name, stand for?

II. ARITHMETIC

The procedure is the same as in the information test, with the one exception that the question is read *twice*. Thus read the number of the question and the question *twice* and then count five silently and pass on to the next question. Let pupils have a rough sheet for figuring.

ARITHMETIC TEST

26. A house rents for $128 for 4 months. How much is the yearly rent?
27. A Ford runs 18 miles on a gallon of gasoline. How far will it run on 5½ gallons?
28. How many pencils can you buy for 60 cents, at the rate of 2 for 5 cents?
29. A boy's suit is bought for 12 dollars and sold for 18 dollars. After allowing 3 dollars for expenses of selling, how many dollars profit?
30. If a train travels half a mile in a minute, what is its rate per hour?
31. What number, multiplied by 7, gives a product of 28?
32. What number doubled is half of eight?

33. A salesman showed a motorcycle at $75. The customer asked to be shown something 20% cheaper. What must its value be?
34. A certain candidate obtained 55% of the votes. There were 500 votes. How many votes did the candidate obtain?
35. If it takes 4 men 3 days to dig a drain 120 ft. long, how many men are needed to dig it in half a day?
36. How many hours will it take a truck to go 42 miles, at the rate of 4 miles an hour?
37. What is the interest due on $400 at 3% for 2 years?
38. A group of girls walks 7 miles east and 3 miles west. How far were they from the place of starting?
39. At the rate of $4 for an 8 hour day, how much is due a man for 32 hours of work?
40. In buying a house for $4,500 I pay 20% down. How much is left to pay?
41. A man had 11 cigars. He smoked 7 and then bought a number, equal to those that were left. How many had he finally?
42. I buy 3 articles, each costing $1.33. I hand the clerk a 10 dollar bill. What change should I get?
43. If an aeroplane goes 250 yards in 10 seconds, how many yards will it go in one-fifth of a second?
44. If there are two tanks, one double the size of the other, and both tanks together hold 12 gallons, how much does the smaller hold?
45. A rectangular plot of ground measures 50 feet one way and 100 feet the other. What length of fencing will go around the plot?
46. What number, when added to 21, makes one-half of 58?
47. How long will it take two trains, just meeting, both travelling at 30 miles per hour, to be 120 miles apart?
48. What number, when squared, is ¼ of 100?
49. If a man is paid 1 dollar an hour for ordinary time, and double for over-time, how much should he be paid for 16 hours, 8 hours being ordinary time and 8 hours over-time?
50. If John gets 4 dollars 30 cents, and Mary gets half of that, how much do they both get together?

III. Opposites

Write the following words on the board:

Up	Asleep
Long	West

and have the pupils give opposites.

Explain to the pupils that you are about to read a series of words, and that each word will be read once and once only. They must not write the word itself but must write the word that means the opposite. This must be done as quickly as possible as only five seconds will elapse between word and word. If they do not think of an opposite, instruct them to draw a dash and wait for the next word. It is well after the explanation is given to get

one or two of the duller pupils to explain just exactly what is to be done.

Read the number of the question and the word; *let five (5) seconds elapse;* and then read the next question number and word, and so forth. Tell the pupils not to get flurried if they miss one item but to wait for the next word.

OPPOSITES TEST

51. summer	60. often	68. sharp
52. slow	61. beginning	69. silent
53. masculine	62. rude	70. expenditure
54. command	63. ancient	71. grant
55. amateur	64. after	72. wild
56. soothing	65. above	73. huge
57. robust	66. come	74. empty
58. innocent	67. wrong	75. depressing
59. heat		

IV. SENTENCE UNDERSTANDING

This is the most difficult test to give as the answer is merely "yes" or "no." In order to insure that the pupils put their answers opposite the right number, the first procedure before explaining the test is to dictate the *last* word of each sentence from 76 to 100, the pupil writing the single word in the place provided. This last word serves in addition to the number, as a mark of identification for the question.

Now explain the general idea of the test with the following samples:

> Are examinations used in educational institutions?
> Is it beneficial to be educated?
> Is perfume a cheap commodity?

When the children understand that they have to answer each question by "yes" or "no," let them return to question 76—the question which has "dollar" written in it—and instruct them to write their answer to the questions one by one as you dictate them, taking care of the numbering. When the pupil does not know or is not sure, tell him to guess. He must write for each question "yes" or "no."

Give the number and the question *once and once only*. Allow five seconds and pass on to the next question. At the eighty-first, eighty-sixth, ninety-first, and ninety-sixth questions warn pupils to follow the numbering carefully.

Test in Understanding Sentences

76. Is a quarter more valuable than a fifth of a dollar?
77. Is a quotient obtained by subtracting?
78. Is secrecy essential to warfare?
79. Do criminals evade justice?
80. Are libraries an asset to a community?
81. Does sarcasm usually cause pain?
82. Is stamp-collecting detrimental to geographical knowledge?
83. Ought the community to eliminate education?
84. Ought charitable institutions to be efficiently managed?
85. Is innocence always established in courts of law?
86. Is tradition eliminated by the reading of historical books?
87. Does patriotism foster anarchy in the state?
88. Is exaggeration essential in scientific writing?
89. Is rapid transportation desirable in large cities?
90. Does exceptional mentality hinder promotion in school?
91. Is competition usually found in recreational activity?
92. Is appetite invariable at all times?
93. Is arithmetical calculation useless in everyday life?
94. Does constancy contribute to amicable understanding?
95. Is revolution a mark of stability in society?
96. Do illustrations usually clarify meaning?
97. Are commercial activities usually on a competitive basis?
98. Is tuberculosis a devastating disease?
99. Is punishment invariably just?
100. Is profiteering a legitimate form of business?

Scoring

Tell the pupils to exchange papers. The examiner or teacher by reference to the key reads out the answer while the pupils score the questions, a tick (✓) for right and a heavy cross (X) for wrong.

It is well to correct items 1-75 and stop there, instructing the pupils to count the number right up to this point. Let this number be x.

Now correct items 76-100. Tell the pupils to count the number right and then the number wrong. The score for the last twenty-five items is obtained by subtracting the number wrong from the number right. This is done to eliminate the guessing factor. Where there are more wrong than right the score for 76-100 is counted as zero. Let the score from 76-100 be y then.

Final score $= x + y$.

one or two of the duller pupils to explain just exactly what is to be done.

Read the number of the question and the word; *let five (5) seconds elapse;* and then read the next question number and word, and so forth. Tell the pupils not to get flurried if they miss one item but to wait for the next word.

OPPOSITES TEST

51. summer	60. often	68. sharp
52. slow	61. beginning	69. silent
53. masculine	62. rude	70. expenditure
54. command	63. ancient	71. grant
55. amateur	64. after	72. wild
56. soothing	65. above	73. huge
57. robust	66. come	74. empty
58. innocent	67. wrong	75. depressing
59. heat		

IV. SENTENCE UNDERSTANDING

This is the most difficult test to give as the answer is merely "yes" or "no." In order to insure that the pupils put their answers opposite the right number, the first procedure before explaining the test is to dictate the *last* word of each sentence from 76 to 100, the pupil writing the single word in the place provided. This last word serves in addition to the number, as a mark of identification for the question.

Now explain the general idea of the test with the following samples:

> Are examinations used in educational institutions?
> Is it beneficial to be educated?
> Is perfume a cheap commodity?

When the children understand that they have to answer each question by "yes" or "no," let them return to question 76—the question which has "dollar" written in it—and instruct them to write their answer to the questions one by one as you dictate them, taking care of the numbering. When the pupil does not know or is not sure, tell him to guess. He must write for each question "yes" or "no."

Give the number and the question *once and once only*. Allow five seconds and pass on to the next question. At the eighty-first, eighty-sixth, ninety-first, and ninety-sixth questions warn pupils to follow the numbering carefully.

Test in Understanding Sentences

76. Is a quarter more valuable than a fifth of a dollar?
77. Is a quotient obtained by subtracting?
78. Is secrecy essential to warfare?
79. Do criminals evade justice?
80. Are libraries an asset to a community?
81. Does sarcasm usually cause pain?
82. Is stamp-collecting detrimental to geographical knowledge?
83. Ought the community to eliminate education?
84. Ought charitable institutions to be efficiently managed?
85. Is innocence always established in courts of law?
86. Is tradition eliminated by the reading of historical books?
87. Does patriotism foster anarchy in the state?
88. Is exaggeration essential in scientific writing?
89. Is rapid transportation desirable in large cities?
90. Does exceptional mentality hinder promotion in school?
91. Is competition usually found in recreational activity?
92. Is appetite invariable at all times?
93. Is arithmetical calculation useless in everyday life?
94. Does constancy contribute to amicable understanding?
95. Is revolution a mark of stability in society?
96. Do illustrations usually clarify meaning?
97. Are commercial activities usually on a competitive basis?
98. Is tuberculosis a devastating disease?
99. Is punishment invariably just?
100. Is profiteering a legitimate form of business?

Scoring

Tell the pupils to exchange papers. The examiner or teacher by reference to the key reads out the answer while the pupils score the questions, a tick (✓) for right and a heavy cross (**X**) for wrong.

It is well to correct items 1-75 and stop there, instructing the pupils to count the number right up to this point. Let this number be x.

Now correct items 76-100. Tell the pupils to count the number right and then the number wrong. The score for the last twenty-five items is obtained by subtracting the number wrong from the number right. This is done to eliminate the guessing factor. Where there are more wrong than right the score for 76-100 is counted as zero. Let the score from 76-100 be y then.

Final score $= x + y$.

For verification at the end of the test collect up information as follows:

Questions	Right	Wrong	Score
1-75	43	32	$43 = x$
76-100	20	5	$15 = y$
		Total Score.....	$58 = x+y$

KEY

INFORMATION	ARITHMETIC	OPPOSITES (*Any reasonable opposite*)
1. July	26. 384	51. summer
2. gasoline	27. 99	52. slow
3. Longfellow	28. 24	53. masculine
4. 8	29. 3	54. command
5. Lee	30. 30	55. amateur
6. Egypt	31. 4	56. soothing
7. deer	32. 2	57. robust
8. Ford	33. 60	58. innocent
9. Thermometer	34. 275	59. heat
10. 19	35. 24	60. often
11. 7	36. 10½	61. beginning
12. 1924	37. 24	62. rude
13. oak	38. 4	63. ancient
14. November	39. 16	64. after
15. Sims	40. 3600	65. above
16. teeth	41. 8	66. come
17. Goliath	42. 6.01	67. wrong
18. shepherdess	43. 5	68. sharp
19. coffee	44. 4	69. silent
20. 98 (98½) (98¾)	45. 300	70. expenditure
21. gum (Spearmint)	46. 8	71. grant
22. CXII	47. 2	72. wild
23. Marconi	48. 5	73. huge
24. alien	49. 24	74. empty
25. Bachelor of Arts	50. 6.45	75. depressing

SENTENCE UNDERSTANDING

76. dollar yes	85. law no	93. life no
77. subtracting no	86. books no	94. understanding yes
78. warfare yes	87. state no	95. society no
79. justice yes	88. writing no	96. meaning yes
80. community yes	89. cities yes	97. basis yes
81. pain yes	90. school no	98. disease yes
82. knowledge no	91. activity yes	99. just no
83. education no	92. times no	100. business no
84. managed yes		

SECTION II

Validity of Test

After trying out the test on several hundred children and making all necessary modifications to bring it to its final form, the hundred questions were administered to two special groups to establish the validity of this method of examination. The first group was a grade VIIA, consisting of thirty pupils of an upper grade VII class, who had been selected from sixty pupils of grade VII by a combination of intelligence tests and teachers' ratings. The thirty pupils therefore formed a fairly homogeneous group, and put a somewhat severe test on an intelligence examination.

The group had previously been given the Otis and Haggerty intelligence tests. Taking the Otis and Haggerty as two typical accepted forms, a comparison of the correlation between these two tests gives a standard of reference as to the reliability of the ordinary form of examination with prepared blanks. Knowing this value and the value of the correlations between the present test and the Haggerty and between the present test and the Otis, an estimate can at once be obtained of the degree of equivalence of the present form of test. The following were the correlations obtained by the rank square difference method.

Seventh grade (upper section) 30 pupils

Correlation between Otis and Haggerty 0.63
Correlation between Chapman and Otis 0.68
Correlation between Chapman and Haggerty . . 0.61

The approximate equality of the three correlations shows that the present form of test is, for this group, certainly as reliable as the less reliable of the Haggerty and the Otis tests either of which, at present, may be accepted as a good average standard.

The second group of 41 pupils was taken at random from approximately 120 children of grade VIII. These 41 pupils when given the Otis and Chapman tests yielded a correlation coefficient of 0.78. Results of this order must necessarily be obtained unless there exists a much higher correlation between intelligence and efficiency as measured by speeded reaction to material read, than that which exists between intelligence and efficiency as measured by rapid oral questions.

The only other point of interest which will be taken up is the range of score for a typical group. As proof of sufficient range a typical grade VII, 108 cases, is distributed in percents.

Score	70	60	50	40	30	20	10	0
Percents	1	3	12	26	26	21	10	1

It will be seen that the range is even greater than that usually given by the Otis, Haggerty or the Chapman and Welles prepared tests.

The merits and demerits of this form of test may be summarized:

ADVANTAGES

1. There is no need for prepared blanks; only pencil and paper are required.
2. The test can be corrected by pupils exchanging papers.
3. Every element of the test is given to each pupil, whereas in the ordinary test the poor pupil only attempts a small fraction of the test.
4. There is a constant stimulation of the subject.
5. It tests an aspect of general intelligence upon which the school is particularly dependent, namely the power to comprehend and follow oral directions and instructions.

DISADVANTAGES

1. A slight lack of objectivity may result due to personal equation of examiner. (NOTE. Even in present prepared forms this factor enters.) In any case the lack of objectivity in the present test is much less than that found in the Binet examination by which the validity of the group test was initially established.
2. There is a possible superiority in the written form of test in so far as it admits of maximum speed for each pupil throughout the whole examination.

Although this test has only been employed for group work with grades VII and VIII and the first year of high school, there is no doubt that it can be used with benefit from grade VI to the third or fourth year of high school.

The author wishes to express his obligation to Mr. Ernest C. Witham and to Mr. W. H. Martin for facilities afforded him in doing this work

SUGGESTIONS FOR PROCEDURE FOLLOWING A TESTING PROGRAM—*I. RECLASSIFICATION*

B. R. BUCKINGHAM
University of Illinois

Two types of procedure seem especially appropriate as a result of testing. One has to do with certain large regroupings of pupils so as to bring under similar influences those who have similar needs and opportunities. This I shall call classification or (emphasizing the idea of change) *re*classification. It is an administrative measure; and, as here used, includes all regroupings of pupils except such as take place within the same class and grade.[1] It therefore includes regroupings which involve changes in any of the following ways: (a) in grade but not in class—e. g., in rural schools; (b) in class but not in grade—e. g., transfers from one section of the seventh grade to another; and (c) in both grade and class—e. g., promotion when the teacher changes. Such regroupings involve the promotion, demotion, and "leaving back" of pupils, their transfer from one section to another in the same grade, and their placement in any of the various kinds of special classes.

The other type of procedure has to do with certain minor regroupings of pupils (in the same grade and class) and with changes in teaching, the object being to subject each individual to specific influences according to his particular needs and opportunities. This type of procedure may be called *individualization*. It is not an administrative but a teaching measure. In the total process of differentiating instruction, reclassification makes the grosser adjustments and individualization the finer adjustments.

Although there is a widespread feeling that the measuring movement cannot be justified unless it leads to better school pro-

[1] I shall use the term "grade" to mean a level or stage of progress in the school. It is to be distinguished from the term "class." The latter will be understood to mean the group of pupils responsible to one teacher. Thus in a given building or school system there may be several classes of the same grade. Moreover, more than one grade may be represented in a single class. Such is the case when pupils of the third and fourth grades are taught by the same teacher. Such is especially the case in the one-room rural school. As to the word "grade" it would be well if educational writers would never use it except to mean "a level of progress in school." Its use as a synonym for *score*, *rating*, or *mark* often causes confusion.

cedure, it is remarkable how little the literature has contained beyond the expression of this feeling. Specific details of what procedures have been followed in actual practice have been exceedingly few. At the close of this article a short bibliography gives what seems to me to be the most helpful material of this nature.

Anyone who consults this material will be struck by its meagerness and by the small measure of success which has attended the efforts of the authors to think the problem through. There are several reasons why this is true—such as the remoteness of most of these writers from contact with pupils, the fact that teachers have not yet been heard from on this topic, the complexity of the data furnished by the tests, and the lack of a suitable method and medium of reporting. But I think the chief reason why neither the school people nor the research workers have given us much on what to do after testing is that test results are expressed in such a variety of ways as to make their interpretation difficult. Units have no important, common significance. Indeed they have almost as many different meanings and sizes as there are different tests. For instance, *one* in the Courtis Arithmetic Test in addition means an example consisting of 27 digits. In the Monroe Silent Reading Test *one* means a word read in a minute. In the Ayres Spelling Scale *one* means a word spelled correctly, while in the Thorndike Handwriting Scale *one* means a certain fraction of the range of ability in handwriting. The diversity involved in these sorts of units is unfortunate.

We may be able to reduce it. First, we may observe that it is not at all necessary that a thing be measured by something like itself. The strength of an electric current need not be expressed as a certain number of times a unit current. It may be measured by anything whose magnitude varies with it—e. g., by the deflection of a needle brought under its influence. Similarly, ability to read need not in theory be measured by rate or comprehension of reading. If quality of handwriting or accuracy in adding varied with ability to read, either could afford a measure of such ability.

Second, we may seek some aspect of school life which varies with all the different abilities which the tests measure. This seems a large task but with certain limitations on the term "varies

with" the quest becomes reasonable. Age, for example, is a factor in school life which, on the whole, varies with school abilities. Again, we consciously seek to register school progress by the grades in which we place children. A large mass of evidence shows that we are doing this very poorly. But despite the shortcomings of the graded system it remains a fact that in general children read better, spell better, and compute better in each successive higher grade. In this sense, therefore, grading varies with the school abilities of children.

Accordingly, I suggest that for all tests *one* be the step from one school grade to the next and that all test scores, no matter what may be their original units, be converted into grade units. This method by affording a common unit at once provides a universal basis of scoring.

Moreover, if, as is apparently the case, research workers and school men believe that *reclassification* of children is one of the most important procedures which may be based on testing, then this form of expressing scores has obvious value. We are not obliged to guess the meaning of a score of 21 on Test II of Monroe's Silent Reading Test nor the meaning of quality 60 on the Ayres Handwriting Scale. We say that the one is 7 and the other 8, meaning that the one corresponds to the beginning of grade seven and the other to the beginning of grade eight. Thus we may transmute these and other scores of different and often obscure meanings into scores having a common, an obvious, and a practical meaning.

To facilitate the conversion of point scores into grade scores Table I is furnished. In this table a few of the tests which have been carefully standardized are represented. The figures, except in a few cases, are not the standards printed in connection with these tests. They represent the best approximation from available standards to the scores corresponding to the *beginning* of each school grade. Sometimes the scores as published represent averages obtained from testing at various times throughout the year. They, therefore, approximate the results which one would expect at the middle of the year. By taking the mean of the published standards for two consecutive years, we may thus secure a standard corresponding to the beginning of the higher of these two years. This has been done in presenting these stand-

TABLE I. GRADE STANDARDS IN VARIOUS TESTS, ADJUSTED TO THE BEGINNING OF THE YEAR

Grade	Arithmetic					Silent Reading				Handwriting, Ayres		Spelling, Ayres (Col. P)	Composition, Willing		Language, Charters (Pronouns)	Geography, Hahn-Lackey (Col. P)
	Courtis Series B (Attempts)				Buckingham Problems	Monroe (Form 1)(a)		Courtis								
	Addn.	Subtr.	Mult.	Divn.		Rate	Comprehension	Rate	Comprehension (Index)	Rate	Quality		Story Value	Form (Errors)		
VIII	11.2	12.2	10.9	10.1	8.2	104 (Test II)	24.5 (Test II)	.....	.....	77	60	96	77.8	4.4	26.7	73
VII	10.3	11.1	9.6	8.9	6.8	96 (Test II)	21.0 (Test II)	191	95	73	56	91	75.3	5.1	22.9	65
VI	9.2	9.8	8.7	7.1	6.3	94 (Test I)	20.8 (Test I)	168	93	67	52	83	71.1	6.3	20.0	54
V	8.0	8.2	6.7	5.3	5.3	79 (Test I)	15.3 (Test I)	145	89	59	48	72	63.4	8.8	16.8	42
IV	6.2	6.2	.....	.....	4.2	60 (Test I)	9.3 (Test I)	113	78	48	44	54	50.1	14.6	14.3	.....
III	.....	.....	.....	.....	3.2	.....	.....	84	59	36	40	31	.....	.....	.....	.....

(a) The scores on Forms 2 and 3 of Monroe's Standardized Silent Reading Test may be reduced to the basis of Form L by multiplying them by the following constants:

	Test I		Test II	
	Rate	Comprehension	Rate	Comprehension
Form 2	0.88	0.95	0.80	0.93
Form 3	0.78	0.94	0.93	0.86

ards in Table I. Sometimes, however, authors have indicated that their standards refer to conditions at the end of the year. Under such circumstances the standard given for the end of the fourth year (for example) has been used as the standard for the beginning of the fifth year and has been entered in Table I opposite Grade V. The essential point is that all the figures purport to show the scores which typical children make or may be expected to make at the time they enter the indicated grades.

For purposes of reclassification, scores appropriate to the beginning of each grade are especially valuable—more valuable, for example, than scores applicable to the middle or end of a grade. When the question of changing the grade of a pupil is under consideration it is frequently decided on the basis of the pupil's supposed ability to *begin* the work of some other grade than the one to which he belongs. Even when "skipping" in the middle of the year or term is being considered, the scores corresponding to the beginning of the year will afford the best point of reference. Suppose, for example, that at the end of the third month of a nine-month school year a series of tests is given including the Courtis Silent Reading Test. One-third of the year having elapsed, we shall be interested in knowing not only the scores appropriate to the beginning of each grade but also those appropriate to each grade after one-third of its work has been done. The scores, when entered as in Table I with reference to the beginning of the year, will best serve the purpose. The beginning of the fourth grade is represented by 113 words per minute and the beginning of the fifth grade by 145. The difference is 32; and one-third of a year after the beginning of the fourth grade (the fourth and one-third grade, so to speak) would bring us to 124 as the best available expression of rate of reading for typical fourth-grade children at the time the test was given. The reader will readily see how much more direct such a procedure is than would be the case if 4 in Table I meant not just 4 but 4.5 or 4.9.

In using Table I interpolation will be necessary. Consider the case of a pupil who scores, on the Courtis Arithmetic test, 9 in addition, 8 in subtraction, 7 in multiplication, and 9 in division. In virtue of scoring 9 in addition he evidently shows ability between the abilities of typical children entering the fifth and sixth grades. Since the score for entering the fifth grade is 8.0 and that

for entering the sixth grade is 9.2, the grade interval is 1.2. This pupil's score is 1.0 above the entry for the fifth grade. He, therefore, shows an ability ten-twelfths greater than that which corresponds to the beginning of grade v. In other words, in grade units, his ability in addition may be represented by 5 10/12 or by 5.8. In a similar way, his score of 8.0 in subtraction may be converted into 4.9 where the grade is used as a unit. His multiplication score of 7 is equivalent to 5.3 on the grade basis; and his division score of 9 is equivalent to 7.1.

Observe also that since we have converted the point scores into scores having the same units, the latter may be combined into an average score. We could not so combine 9 addition examples with 8 subtraction examples. Even more clearly we could not combine 9 addition examples with 6 errors in composition or 75 letters per minute in handwriting. But grade scores in as many tests as are provided with grade standards may be amalgamated into a general grade score. As I have just indicated, the pupil whom we have been considering had the following grade scores: addition 5.8, subtraction 4.9, multiplication 5.3, and division 7.1. His average grade for these four items was therefore 5.8. In general, he may be said to have performed nearly as well as typical pupils entering the sixth grade.

Some point scores are of such a nature that small abilities are shown by large numbers and large abilities by small numbers. In other words, the smaller the score the better the performance. This condition arises, for example, when the score expresses the time required to do a given amount of work or the number of errors made in doing it. Such a series of scores is given in Table I for errors of form in composition. Without some means for bringing expressions of this sort into line with those in which larger numbers indicate greater ability, we are in difficulties. The method of converting point scores into grade scores provides such a means. For example, a score of 8 formal errors in composition (within the meaning of the Willing scale) yields a grade score of 5.3.

In reducing point scores to scores in terms of grades we are not only securing a uniform basis but one in which the units are significant for the purpose we have in mind. While it is true that the increase in ability from one grade to the next is not the same throughout the elementary school, these increases are

equal in the vital sense that they are acquired in equal lengths of time. In order that a numerical expression may be reliable and useful, the units of which it is an aggregate need not be the same in every sense of the word. It is sufficient if they are the same in one sense only, provided that is the sense in which the measures are understood and used. For purposes of administrative procedure the grade as a unit is decidedly preferable both on theoretical and practical grounds to the various units in which point scores are expressed.

It is clear from what has been said that for reclassifying pupils educational (i. e., subject-matter) tests may be used to advantage. Intelligence tests, however, provide the most important type of information supplementary to that obtained through educational testing. Indeed, intelligence tests have been widely and usefully employed without reference to educational tests; and they have been used with particular reference to reclassification. Authors of them have avoided one of the mistakes of authors of educational tests. They have not used point scores (which have different meanings for different tests) but a common unit, namely, the mental age. With knowledge of the age appropriate to each grade, users of intelligence tests can also convert mental ages into grade scores.

At the conclusion of this article a few titles are given which will guide the reader to some of the best material on the use of intelligence tests for classification purposes. In at least one instance which has come to our attention[2]—and probably in others which have not—the children in an entire school were distributed according to their mental ages and according to their grades at the time of testing. This was done as a preliminary step in an effort to decrease the number of children whose mentalities were either above or below the level appropriate to their grades. The report indicates what can be done by organizing special groups of children, utilizing coaching teachers, and being generally watchful to bring children into their appropriate grades.

Despite the general belief in the efficacy of intelligence tests as a basis for classification, they seem to me to be insufficient for

[2] Pintner, Rudolf and Noble, Helen. "The Classification of School Children According to Mental Age," *Journal of Educational Research*, 2:713–28, November, 1920.

this purpose. No doubt a child whose mentality is six or six and a half can safely be placed in the first grade. No previous instruction is required. But whether a child whose mental age is seven and a half can at once begin the second grade, depends on what instruction he has received. He certainly cannot do so if he has had no instruction. In general it seems reasonable that when pupils evince a mentality corresponding to a higher grade than the one in which they are located, some evidence of preparation for the higher grade should be obtained before advancing them to it. *This evidence it is precisely the function of the educational tests to afford.*

An example will illustrate. Suppose a third-grade child has shown according to an intelligence test a mental age of ten years. This indicates so far forth that he should be doing fourth-grade work. If, however, tests in reading, arithmetic, language, etc., do not indicate a performance on a par with his mentality, we are justified in hesitating to classify him in the fourth grade. If, on the other hand, his performance in these subject-matter tests is at or above par, we are strengthened in our belief that he can do the work of the fourth grade. We have evidence not only that he possesses the required mental ability but also that he can put his mental ability to effective use.

Again, a mental test alone may indicate that a sixth-grade pupil has only fifth-grade ability. Nonintellective qualities, however, may be favorable in his case. Educational tests, if given, may show that he is actually meeting or surpassing the requirements of his grade. We have abundant evidence that such cases exist. It is clear that reliance upon intelligence tests alone—a reliance which would lead us either to demote or to leave back the pupil whom we have just described—would be unfortunate.

On the other hand, despite the attractiveness of the grading scheme presented in connection with Table I above, testing on subject matter affords only a partial guide in the reclassification of children. The fact that a fourth-grade child shows fifth-grade ability in arithmetic does not give us conclusive evidence that we ought to teach him arithmetic in the fifth grade. If we find that his mental ability is clearly better than is indicated by a mental age of eleven (the fifth-grade age), we may infer that with a little preliminary instruction he should be placed in the

sixth grade. This will be especially true if he has had no instruction hitherto in either fifth- or sixth-grade work.

On the contrary, even though a pupil's performance may equal that of the average fifth grade, a knowledge of his mentality may lead us to believe that he will not succeed in the fifth grade. He may be a child of fourteen with a mental age of eleven. His intelligence quotient or rate of development would then be but 79 percent of normal. If such a child were in the fourth grade, he would probably be unable to do satisfactory fifth-grade work in arithmetic despite the fact that his performance at the time of testing might equal that of the average fifth-grade child.

But the insufficiency of subject-matter testing is most apparent when we consider that it offers no evidence of whether pupils are performing as well as they should—no evidence of whether the school has been accepting poorer work than it would have been justified in demanding if a knowledge of the mentality of the pupils had been available.

Thus, neither educational testing alone nor intelligence testing alone is complete. As a basis of classification we desire to know both what children might do and what they have actually done. We recognize as factors in their success other personal qualities than intelligence; and we recognize that the work of the school as it influences them varies in amount and quality. These personal and institutional characteristics we can scarcely isolate, nor can we subject them to separate measurement in the case of each child. This, however, does not justify us in ignoring them. If we cannot measure each of them independently, we can at least measure their combined effect. This is exactly what we do by means of educational tests.

The result, however, as thus obtained is a record of the effect not only of the nonintellective (personal and institutional) factors but also of the factor of intelligence; and it is the effect of all these factors combined. But intelligence (which is essentially the ability to learn) is so important that we have rightly sought and obtained an independent measure of it. After obtaining it in the case of an individual child and after securing his performance on a series of educational tests, we know two things neither of which may be neglected. First, we have a measure of the one most important factor in his school success—

intelligence; and from this we are entitled to infer what might have been done, if other factors had been normally operative. Second, we have a measure of the pupil's school success itself. It may not be as great as his intelligence would have led us to expect, or it may surpass such expectations. In either case it tells us something about the operation, favorable or unfavorable, of all the other factors.

It is only when in some way these two sorts of data are combined that we may arrive at a defensible plan of action with reference to the pupil in question. We may do this on the assumption that the nonintellective factors are going to continue to operate as they have in the past; or we may plan our action on the assumption that certain of these factors will operate more favorably in the future. In particular, we may resolve that the institution will itself play a better part. A quickened sense of responsibility and of opportunity among teachers and pupils may manifest itself; and we may justifiably infer from the intelligence tests that the grade a pupil might have reached is also the grade which under improved conditions he may yet reach.

Clearly, what we need in this connection is a method of bringing intelligence and performance together. Since we recognize that, due to characteristics other than intellectual, a pupil's performance may be above or below that which his intelligence leads us to expect, we shall want some measure of the discrepancy. The educational quotient or achievement quotient has been offered as such a measure. This means that the measures of intelligence and of performance must be expressed in the same units. They may be expressed in years (ages), as is done in the battery of tests known as the Illinois Examination; or they may be expressed in school grades as is suggested above. In either case the extent to which performance exceeds or falls short of mentality may be at once found either as a difference or—with some added advantage—as a quotient. This difference or quotient is an expression of the extent to which, as elicited by the school, the factors which I have called nonintellective have exerted their influence. More accurately it measures the extent to which such factors have been above or below their normal potency. Such measures are highly important. Apart from their use in reclassification, they have decided value as indicating the extent to which

the school is making the most of its pupil-material. Obviously, however, these advantages cannot be secured unless both intelligence and educational tests are used.

In matters of reclassification, graded schools and ungraded schools[3] will proceed somewhat differently. Each type of organization has certain advantages and disadvantages. Each should make the most of its advantages.

In the graded school there is, or should be, opportunity for the organization of classes having specific aims and consisting of pupils of similar abilities. Superintendents, principals, and teachers should cooperate in organizing groups of children, whether under separate teachers or not, which have definite purposes. These purposes should be known to all—to pupils, teachers, and officials. Under these circumstances a great deal of work can be done. If a sixth-grade teacher has six pupils who according to mentality should be in the seventh grade and four who ought to be in the eighth grade, what she should do is to acquaint each of these groups of children with the object in view and give them extra work, added attention, and plenty of encouragement to make up the loss which the school has entailed for them through ignorance of their capacities. Educational tests will first indicate how great the loss is and will afterwards record the extent to which it has been recovered.

Teachers will not do this kind of work without an incentive. Naturally, they like to keep their bright pupils. A principal or superintendent would, in my judgment, do well to make much of every additional promotion of a child apparently qualified for a higher grade. Such an official would be justified in publishing such facts and in making them the subject of personal commendation. In what I have been saying I have been thinking more of bright than of dull children. There are indications that there are quite as many children whose mentality is above their present status as there are whose mentality is below it. Moreover, the school has paid less attention to bright than to dull children; and, as has often been pointed out, the instruction of bright children pays large dividends to society.

[3] By the term "ungraded schools" is meant particularly the one-room rural school. To a certain extent, however, it includes partially graded schools—i.e., those in which more than one yearly grade is taught by one teacher.

Before leaving this topic, I should like to say that in applying remedies, as the need for them is disclosed by testing, I am no advocate of precipitate action. If a fifth-grade child shows seventh-grade mentality, he should not, in my judgment, be placed in the seventh grade until intermediate work has been to some extent covered. All that I should urge is that the seventh grade should constitute a goal or objective toward which both the teacher and the pupil should work. Of course, as time passes this goal will recede; and by the time another year has elapsed it will no longer be the seventh but the eighth grade which should be attempted.

As to dull children—children who score low in both intelligence and educational tests—logic would suggest that we make the appropriate grade an objective as in the case of bright children. But here logic will not meet the need. It would require us to demote backward children as freely as we should promote gifted children. This we hesitate to do. Moreover, it is a question whether it is desirable from a social point of view. And we certainly cannot expect a pupil to work to secure a reduced grade. There is no motive and no hope of the pupil's cooperation, as there is when a higher grade is proposed. Perhaps we should simply fail to promote these children at the end of the promotion period. The best provision, however, for dull children—best for them and best for others—will consist in the organization of special classes.

The ungraded school is at a disadvantage in many ways; but in the ease with which children may be reclassified, it is in a more advantageous position than the graded school. Suppose a child of the fourth grade is found to have a mental age of eleven years. Since eleven years corresponds to the fifth grade, this means, as has been indicated before, that if he had been properly taught from the beginning, he would now probably be in that grade. Results of subject-matter testing may corroborate the results of mental testing; or, in less favorable cases, the former may be lower than the latter. In either event, if the child is made acquainted with the fact that he can do the work of the fifth grade provided he gets ready for it, the presence of the fifth grade reciting in his hearing and to the same teacher creates a natural path along which his energies may be directed. Not only can he hear the fifth grade recite, but he can also participate with it as he could not if

he were in a graded school. From the very nature of the case there can be no conflict of fourth- and fifth-grade recitations. Both classes recite to the same teacher and hence of necessity at different periods of the day. If the teacher plans to have the child of whom we are speaking move forward into fifth-grade arithmetic, she can actually have him recite with both the fourth grade in which he now belongs and the fifth grade which is set up as a goal. Material in the course of study between these two grades can be taken up, so far as it is necessary to do so, with very little expenditure of time.

This last statement is not a glossing over of difficulties. To be sure it is logical to suppose that the child who abruptly moves into the fifth grade from the fourth will be handicapped by the loss of intervening subject-matter. But in reality the logic of the situation again fails to work. There is abundant evidence that the handicap through the supposed loss of this intervening material is not as great as we should expect. Even in arithmetic where we should first look for these "gaps" in a child's knowledge, they seem to be imperceptibly closed; and this appears to be true even when little attention is given to them. Not only is the published evidence on the question convincing but the experience of school people also supports it. Many a teacher who has "skipped" a child —sometimes with misgivings—has been obliged to concede that the child in question made good in every observable way.

Opinions differ as to the extent to which reclassification should take place. With reference to superior children the shades of belief are particularly numerous. They range from a sole reliance upon reclassification in the form of extra promotion to a sole reliance upon changing the course of study. I have been directing the reader's attention especially to reclassification.

It has also been my purpose to suggest a method of interpreting test results which will have obvious bearing on reclassification, and to point out the advantages in the same connection of a testing program which includes both intelligence and educational tests. Yet in a certain very real sense we do not reach the heart of the problem of using test results until we take up the question of the finer adjustments—the teaching processes—which I have covered by the term Individualization. I propose to devote a subsequent paper to procedure of this type.

BIBLIOGRAPHY

A. Based on Educational (Subject-Matter) Tests

(1) Gray, William S., "The use of tests in improving instruction," *Elementary School Journal*, 19:121–42, October, 1918.

Suggestive, but necessarily vague as a guide to practice. Recommends systematic word study, and special exercises without saying what ones were used.

(2) Kallom, Arthur W., "The importance of diagnosis in educational measurement," *Journal of Educational Psychology*, 10:1–12, January, 1919.

Mentions some concrete procedures and contains some shrewd advice.

(3) Uhl, W. L., "The use of the results of reading tests as bases for planning remedial measures," *Elementary School Journal*, 17:266–75, December, 1916.

Contains some concrete data on what was actually done. The necessary drudgery incident to an analysis of the question papers had evidently been undertaken.

(4) Barton, J. Wesley, "School organization on an objective basis," *Educational Administration and Supervision*, 6:187–97, April, 1920.

Used Kansas Silent Reading, Courtis Arithmetic, and a spelling test, as well as an opposites test.

(5) Ruch, G. M., "An experiment with forced promotion," *Educational Administration and Supervision*, 6:71–73, February, 1920.

Uses Courtis Arithmetic, Kansas Silent Reading, and Gregory Language.

(6) Henmon, V. A. C., "Improvement in school subjects throughout the school year," *Journal of Educational Research*, 1:81–95, February, 1920.

Shows that advantage may be derived from a testing program without doing anything further than administer the test. Concludes that attention to record-making is, in and of itself, an efficient cause of improvement.

(7) Brooks, Samuel S., "Putting standardized tests to practical use in rural schools," *Journal of Educational Research*, 1:392–97, May, 1920; 2:425–35, June, 1920; also the November, 1920 and January, 1921 numbers of the same journal.

The last two are particularly helpful. Grade scores rather than point scores are used.

(8) Willing, Matthew H., "The encouragement of individual instruction by means of standardized tests," *Journal of Educational Research*, 1:193–98, March, 1920.

The value of this brief article lies in its analysis of errors.

(9) Richardson, J. W., "The campaign method in elementary education," *Journal of Educational Research*, 2:481–92, June, 1920.

The method described consisted essentially in an initial test, a period of intensive effort, and a final test. Large improvements were recorded.

(10) Monroe, Walter S., "Improvement of instruction through the use of educational tests," *Journal of Educational Research*, 1:96–102, February, 1920.

General considerations with reference to class results. Specific procedures not given.

(11) Certain of the city school reports are decidedly important for data on measures taken as a result of testing. The following are worth noting:

(a) *The eighth report of the board of education of Louisville, Kentucky* (July 1, 1918 to June 30, 1919).

(b) *Sixty-fourth annual report of the school district of Ann Arbor, Michigan,* (1920–1921).

(c) *Report, board of education, Duluth, Minnesota, 1918–1919.*

B. Based on Intelligence Tests

(1) Whipple, Guy M., "The problem of selecting and training gifted children in the public schools," Report of the Fifth Conference on Educational Measurements, pp. 6–16.

How the class for gifted children was organized and conducted at Urbana, Illinois.

(2) Arthur, Grace, "An application of intelligence tests to the problem of school retardation," *School and Society*, 10:614–20, November 22, 1919.

Reports an intelligent and consistent use of mental tests to reduce retardation.

(3) Frasier, George W., "The measurement of intelligence as an aid to administration," *Educational Administration and Supervision*, 6:361–66, September, 1920.

Gives specific data in a series of brief case reports.

(4) Terman, Lewis M., "The use of intelligence tests in the grading of school children," *Journal of Educational Research*, 1:20–32, January, 1920.

Classification as a major outcome of testing is set forth as completely and convincingly as the scope of the article permits.

WHAT SHOULD TEACHER-RATING SCHEMES SEEK TO MEASURE?

RAYMOND A. KENT
Superintendent of City Schools, Duluth, Minnesota

For several years past there have been numerous attempts on the part of school administrators and supervisors to formulate a satisfactory scheme of rating teachers. Examination of the various plans as they appear, as well as experience in using some of them, convinces the writer that there is a fundamental weakness or deficiency in practically all schemes of this kind up to date. A careful consideration of some factors involved will help to make the issue clear.

A teacher-rating device in administration or supervision is a procedure for arriving at a judgment of the value of an individual instructor in the school industry. The real basis for judging this value is the output the individual can show as a result of his labor. This does not mean that the sole criterion of deciding on the retention, promotion, or demotion of a teacher is the progress objectively measured which her pupils make under her instruction within a given time. Such a criterion is only a partial one. All the teacher's work, including every major factor of it, should be considered; but these factors should be considered only with respect to what they contribute toward educational results in the children under her care.

Thus far, in formal rating schemes the composite resultant of teaching and the effort of the teacher have received either minor consideration or none at all. This is so anomalous a condition that the assertion may be doubted. That it exists, however, is proved by an examination of individual teacher-rating schemes. For instance, the score card devised and used by Ohio high-school inspectors, as reproduced in the May-June issue of *Educational Administration and Supervision*, does not include the mention of educational product or result. It is quite possible that the omission was intentional. The December issue of a current educational monthly which lies before us as we write contains the latest teacher-rating card we have seen. To "teaching power" it gives 25 percent; while "results" (being one of ten subdivisions judged of mutually equal value) is weighted 2½ percent! A recent

careful examination of a half-dozen or more teacher-rating schemes not only failed to find one where either "results" or "product" was given as much as 40 points out of a total hundred but also an entire lack of any central tendency to make the teaching product the basis of determining the value of the teacher.

There are three possible excuses for this condition. One is that these rating schemes, instead of attempting to formulate a basis for judging the value of a worker as such, have in reality tried to formulate a basis for judging the worker's possession of the characteristics necessary for success.

If this is the function of these devices, then let us frankly say so and limit their use largely to judging individuals before they are formally admitted to the profession. To judge whether a person possesses the qualities necessary for success in any profession or industry is vocational guidance. It is true that such guidance can well continue after the individual has entered the profession. But it is utterly fallacious to suppose that the determination of the existence of potential abilities, even if such a determination is scientifically made, is synonymous with measuring actual working efficiency. We must get away from the idea that because a teacher has a pleasing personality, good training, a cooperative spirit, and ability to write a good lesson plan, she will make a first-class instructor. We must get clearly the idea that no one is a first-class instructor whose training, work, and school associations cannot produce the results which society legitimately demands from the public school system. The fact is that there has been no discrimination made between score cards for rating student teachers and those for judging the merit of a worker in the profession. It is possible there should be none. If the latter be true the emphasis must be shifted. We must not continue to apply to skilled workers the gauge of potentiality. We should rather begin to apply to the apprentice the measure of productive ability.[1]

Another possible excuse is that one may not believe in the possibility of measuring the results of teaching. It is doubtless true

[1] Since the above was written, the writer has examined the score card published by L. B. Hill in the 1920 report of the Society of College Teachers of Education. This card is "for rating practice teachers and teachers in service." The item "skill in teaching process" is the only one listed that connotes results. It may receive a maximum weighting of 20 percent. But not all of its five subdivisions definitely measure the *school product*. There is a very definite emphasis on the *process of teaching* rather than on the *results of learning*.

that many who use the rating systems already formulated, and possibly some of those whose judgments entered into the determination of the Boyce Scale and similar ones, do not believe that the work or the product of the school can be measured objectively. To the degree that this condition obtains it is unsound procedure to include the judgment of these individuals in the formulation of a scale for rating teachers. On the other hand, if one does believe that objective measurement is possible and desirable for school work, then to accept any scheme for judging the individual worker which does not lay the larger emphasis upon the measurable product of that worker is to make our practice inconsistent with our theory.

A third possible excuse is concerned with the idea that the means for making quantitative and qualitative measures of the school product are at present too limited to provide an adequate basis for judgment. It must be admitted that such an idea contains an unquestionable element of truth. It is impossible at present to measure an educational worker's output as rigorously or as inclusively as that of an industrial worker. But a good beginning has been made. If we accept as the objectives of education, pupil achievement in acquiring useful knowledge, applied skills and proper ideals, the problem becomes clearer. We are already able to measure objectively the achievement of children in knowledge, and to a limited extent in skills; but we have no such measure of product in ideals. It is true that before any satisfactory or scientifically sound rating can be made, we must know what are the qualitatively and quantitatively measured products which the teacher secures. But it is also true that the absence or the limitation of these measures should not blind or confuse us as to the real issue involved in the problem of measuring the value of an educational worker. It should be obvious that schemes of teacher rating which give from 0 to 40 percent as the maximum weighting for the product secured by the worker are in this respect clear violations of sound principles in educational administration and supervision. Products or results should be given greater weight than any other single item. We believe that they should receive more weight than all the rest of the factors considered in judging the value of a teacher in service.

A good illustration of the type of a score card which we should strive for is the pupil's report card used at Public School 33, The

Bronx, and described in *Educational Administration and Supervision* for November, 1919. A child's ability in reading is divided between thought and expression, the first having a maximum of fifteen points and the second a maximum of ten. Grammar, word study, and arithmetic are each treated in a comparable manner.

A consideration of certain items in their bearing upon the problem of teacher rating as here discussed may help to clear up some troublesome questions. Suppose for example, that the single factor of morality is given a maximum weighting of fifteen out of a total of one hundred points. Suppose further that a teacher who receives seventy-five or more points may retain her position. It would be entirely possible theoretically for a teacher to receive but five points in morality and still be given a total of seventy-five or more. Yet what school administrator could justify his action in retaining a teacher who justly received only one-third of the possible points assigned to morality? The logical use of the rating plan would compel the retention of such an individual. Sound practice would expel her. Just why should she not be retained? For the simple reason that as a teacher with such limited moral assets, her school product—at least in ideals—would fall far below what it should be. The illustration taken is not unfair to present rating schemes. A careful examination of them will disclose a similar weakness in connection with the weighting of many other factors. The true significance of certain factors which are given small weight in most rating schemes only becomes apparent when regarded from the point of view of results. So regarded the absence of these factors or their presence in low degree is found to make the difference it should make and which in actual practice it does make.

Again, why do we wish a teacher to have a good personality? Simply because personality is a factor which functions toward the achievement of worthy results. Many teachers who possess mediocre personality are found after a period of service to merit a high rating on the basis of their ability to secure desirable school results—a higher rating even than those whose personality receives all the points possible to be accredited to it, but who are lacking in some other respects.

Consider the item of the teacher's effort. How much should it count in the final rating? Shouldn't one who exercises this

meritorius characteristic therefore receive points additional to those given a worker who shows less effort? There is apparent merit in this contention; for effort will tend to capitalize potential ability. A gifted person may not perform as well as one of mediocre endowment because of a difference in effort. On the other hand, absence of native endowment can never be atoned for by sheer effort. It is therefore fallacious to credit effort alone. The true measure is the product which individual ability plus effort is able to bring forth.

Another phase of the problem should also be kept in mind. The accurate measure of any worker's ability can be taken in terms of product only when the quality of the raw material is known. The work of two teachers of equal effort and equal ability may show entirely unequal products qualitatively and quantitatively. This may be due to differences in the intelligence, in previous training, in physical condition, or in home environment between the groups of children with whom they are working. Indeed, the same worker may accomplish less in measured product one year than another and at the same time expend even greater effort to secure the more meager results. But this consideration need be no real obstacle to the use of results as the sound basis for judging teaching merit. By means of intelligence tests we are able to measure the native ability, and by means of standardized tests the acquired ability, of a group of children. Thus we are able to arrive at some conclusion concerning the quality of the raw product which any teacher handles as a professional worker.

If present devices for determining a teacher's merit are unsatisfactory, what device is better? The following plan is submitted as a partial answer to this question. It is not proposed as final or complete, but rather as suggestive and illustrative of the principle for which we have argued in this discussion.

MEASURING A TEACHER'S MERIT AS A PROFESSIONAL WORKER

I. Pupil Achievement—Minimum 50, Maximum 75
   1. Objectively measured (for each subject)
      A. Knowledge
      B. Skills

2. Other Achievements
   A. Habits of Study
   B. Attitude toward
      (1) Work
      (2) School government
      (3) School organizations
      (4) Moral questions
      (5) Life preparation

II. Merit in Mechanics of Worker—Minimum 20, Maximum 40
   1. Organization and administration of
      A. Tools
      B. Raw materials
   2. Skill in Technical Method

III. Merit as a Social Worker—Minimum 20, Maximum 40
   1. Cooperation with organization
   2. Professional habits
   3. Success in dealing with parents
   4. Sympathetic interpretation of pupils
   5. Discharge of obligations as community member

In the administration of any such a plan for rating the following would need to be taken for granted:

1. The knowledge of the potential ability of pupils to achieve, measured in terms of their intelligence.

2. A statement of pupil achievement at the beginning of any period over which the efficiency of a teacher's work is to be judged.

3. The working-out of intelligible standards in the items to be listed under "other achievements."

4. Specific statement by the supervisor or in the course of study, or both, of the tools and methods to be used and how to use them.

5. Statement of intelligible and reasonable standards under which several factors combine to make a teacher successful as far as she is a worker in a social institution.

GRADE NORMS FOR THE NEW YORK CITY PENMANSHIP SCALE

EUGENE A. NIFENECKER

Director of Bureau of Reference, Research and Statistics, Department of Education, New York

In February, 1915, the New York Board of Education adopted a uniform style of letters to be taught in all schools and authorized the teaching of the muscular-movement method of writing which had been tried out in a number of schools for several years.

The next step taken was the construction of a scale from the product of this system of penmanship. This was undertaken by Messrs. Lister and Myers, instructors in the Brooklyn Training School for Teachers. It was adopted by the Board of Education in February, 1918.

The setting up of grade norms for each element measured by the scale was undertaken by the Bureau of Reference, Research, and Statistics in June, 1919, and is the subject of this report.

THE NEW YORK CITY PENMANSHIP SCALE

The New York City Penmanship Scale possesses certain characteristics which distinguish it from the scales that preceded it. In the first place it is a scale constructed from the product of the system of penmanship it is intended to measure. In the standardization of any educational product it is of fundamental importance to know "that the pupils have received the instruction in the field of knowledge covered by the scale or the test. Unless scales or the standards derived from standard tests are based on the results achieved by children after proper instruction, then standards established on the basis of the results from such tests are unsatisfactory as measures of instruction." (Ballou)

The second characteristic of the scale is the fact that it is analytic. In judging ability in penmanship there are so many things to be considered that a scale based only on general merit is of little assistance to the average teacher. The scale should help the teacher to look for the essentials in judging a given specimen of writing. Legibility, ease in writing or freedom of movement, good letter formation, uniformity of size, uniformity of slant, clear

and uniform spacing, neatness, and an easy, fluent appearance—all these make up the value of writing. The authors of this scale considered it possible and practicable to group most of these qualities under three headings.

Accordingly, as they say: "Under the caption *Form*, it was decided to consider accuracy in letter formation, with the standard letter forms adopted for the schools of New York City as the basis, uniformity of size, and regularity of correct slant.

"Under *Movement*, the specimens were judged as to quality of line. As is well known to students of penmanship, heavy, tremulous, and broken lines are evidence of either finger movement or muscular movement that is too slow and deliberate to produce skill; and clear-cut, smooth lines are evidence of the use of the easy-flowing muscular movement that is desired.

"Under *Spacing* the writing was judged as to correct and uniform spacing between letters, between parts of letters, and between words."[1]

As to the construction of the scale, the authors say in the same article: "The scale represents the average judgment of 21 teachers and penmen, experts in the muscular system of penmanship, and 4 psychologists. From 9 schools of Greater New York, representative of the best, medium, and poorest product of the muscular system of penmanship, 3,550 specimens were selected from at least one entire class of each grade from 3B to 8B inclusive."

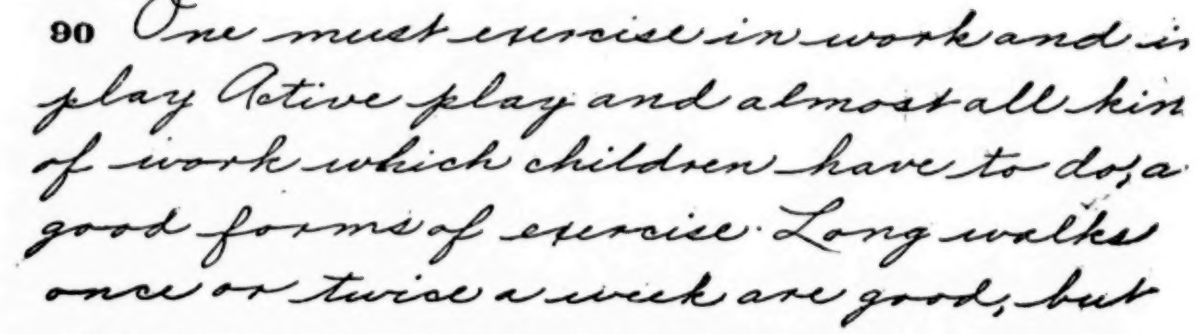
Figure 1 illustrates the resulting scale. The adjacent specimens are two steps apart.

Figure 1. The Scale Illustrated

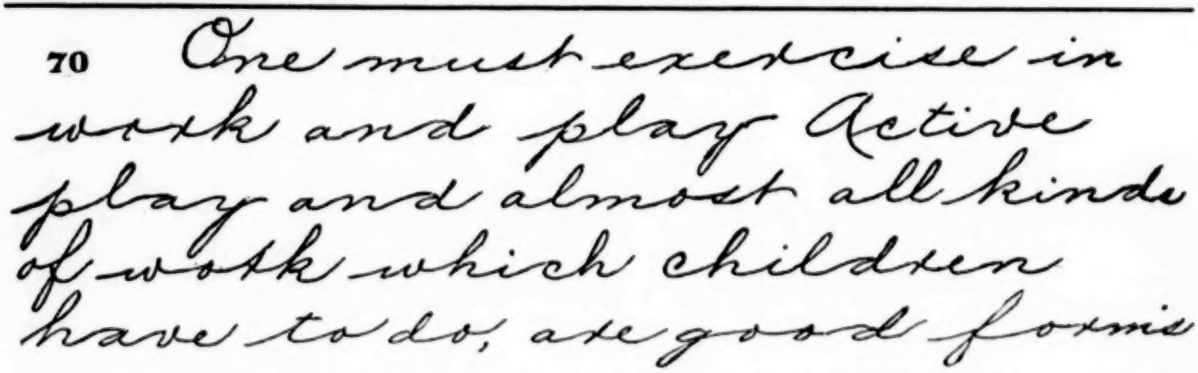
The scale consists of twenty-four specimens arranged in parallel columns—eight under Form, eight under Movement, and eight under Spacing. The scale values of the specimens under each element are 90, 80, 70 . . . 20. Four specimens under each element are reproduced below. These represent alternate steps beginning with 90.

[1] Lister, Clyde C., *The New York City penmanship scale* (Bulletin of the Brooklyn Training School for Teachers, no. 3, January, 1919, p. 4).

FIGURE 1. FORM


90 One must exercise in work and in
play Active play and almost all kin
of work which children have to do, a
good forms of exercise. Long walks
once or twice a week are good, but

This is good elementary school penmanship. Note the uniformity of size, slant, and alignment. The x is crossed carelessly.



Good form. The curvature between letters is slightly exaggerated. Note the approach to a, d, and m.

Good form. The curvature between letters is slightly exaggerated. Note the approach to a, d, and m.

50 One must exercise in work and in play. Active play and almost all kinds of work which children have to do, are good forms of exercise. Long

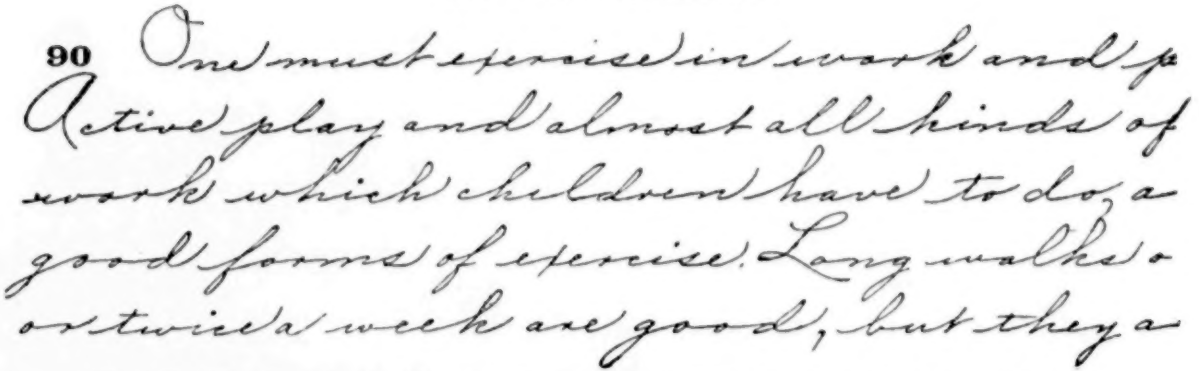
This writing slants too much. The t should not be looped. Note the careless tendency in completing k.

---

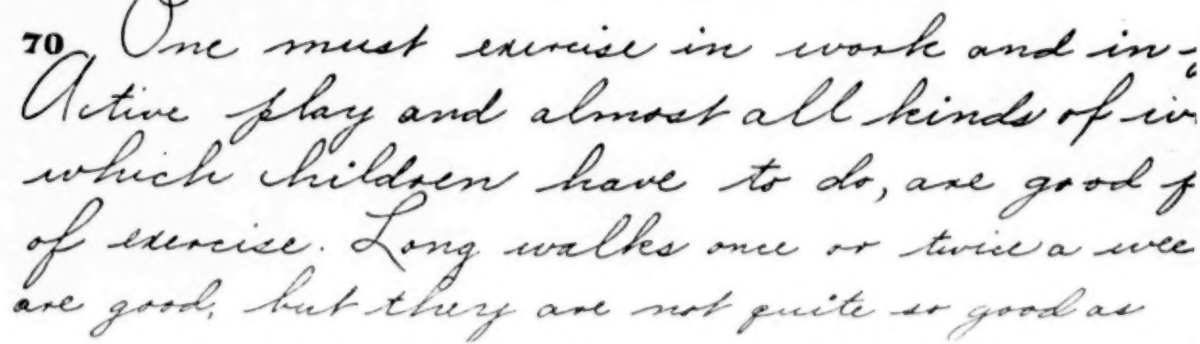
30 One must exercise in worke and in play. Active play and almost all kinds of work which children ha to do, are good forms of

This writing slants too much. The loops are too long and narrow. The t should not be looped. The a and d are poor.

FIGURE 1. MOVEMENT


90
One must exercise in work and p
Active play and almost all kinds of
work which children have to do, a
good forms of exercise. Long walks o
or twice a week are good, but they a

The quality of line shown above is evidence of an easy muscular movement, which is produced by the muscles of the arm.


70
One must exercise in work and in-
Active play and almost all kinds of in
which children have to do, are good f
of exercise. Long walks once or twice a wee
are good, but they are not quite so good as

This shows good movement and fairly good control throughout. Note the sharp, clear-cut quality of line.

This shows good movement and fairly good control throughout. Note the sharp, clear-cut quality of line.

50 One must exercise in work and play. Active play and almost all kinds of work which children have to do, are good forms of exercise. Long walks once or twice a

The smooth and uniform quality of line indicates fairly good movement. The n, r, s, and k are poorly formed.

30 One must exercise in work and in play. Active play and almost all kinds of work which children have to do, are good forms of exercise.

Long walks once or twice a week

Although the movement is poor, this writing shows a better quality of line than the specimen below.

FIGURE 1. SPACING

90 One must exercise in work and play. C
play and almost all kinds of work which
children have to do, are good forms of exer
Long walks once or twice a week are good;
they are not quite so good as regular exer

This specimen shows a high degree of excellence in spacing. The writing, however, is too small.

---

70 One must exercise in work and in
play. Active play and almost all
kinds of work which children
have to do are good forms of exer
cise. Long walks once or twice

This spacing is quite clear, but it is too variable. Compare One, kinds, and once.

This spacing is quite clear, but it is too variable. Compare One, kinds, and once.

50 One must exercise in work
and in play. Active play and
almost all kinds of work
which children have to do
are good forms of exercise.

This spacing is clear, but variable both within and between the words. The slant is variable.

---

30 active play and all mu
all kinds of work which chu
have to do, are good forms o
exercise. Long walks once o
twice a week are good, but t

The separations between letters and between words are not clear. Some lines look like a long word.

TABLE I. AVERAGE DEVIATIONS PER SPECIMEN FROM "EXPERT" SCORES FOR 86 PROSPECTIVE JUDGES—PRACTICE SAMPLES

Element	Average Deviations																	Average
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Form	1	.....	.....	4	8	16	12	25	14	2	1	1	.....	.....	1	.....	1	7.9
Movement	.....	.....	.....	.....	3	8	3	14	15	15	17	3	3	1	1	1	2	10.1
Spacing	.....	.....	1	2	11	24	20	18	1	5	2	2	.....	.....	.....	.....	.....	7.4
Total	1	.....	1	6	22	48	35	57	30	22	20	6	3	1	2	1	3	8.3

TABLE II. DEVIATION ACCORDING TO AMOUNT

Deviations	Form		Movement		Spacing		Total	
	No.	Percent	No.	Percent	No.	Percent	No.	Percent
Zero deviations	300	26.9	171	15.2	216	18.0	687	20.0
5 points (½ step)	288	25.9	374	33.2	543	45.3	1,205	35.1
10 points (1 step)	290	26.0	239	21.2	215	18.0	744	21.6
15 points (1½ steps)	124	11.1	136	12.1	162	13.5	422	12.3
20 points (2 steps)	112	10.0	205	18.2	63	5.2	380	11.0
Total	1,114	99.9	1,125	99.9	1,199	100.0	3,438	100.0

The specimens were divided into four piles according to general merit, and 300 were then selected at random, about the same number being taken from each pile. The 25 judges arranged them in eight piles on the basis of equal intervals in merit. Each judge ranked the specimens three times—namely, as to form, movement, and spacing.

The next procedure is described as follows: "On the basis of the average rank assigned each specimen, the best and the poorest were selected as the top and the bottom of the scale. Therefrom the exact numerical rank which the other six samples should have was determined. The specimens whose average ranks are the same as these determined positions, or are nearest them were selected. Without exception all the samples on the scale are less than .1 from the *determined* position."[2]

Measuring Achievement in Penmanship

Conduct of tests.—In the penmanship survey undertaken by the Bureau of Reference, Research and Statistics, particular care was taken to secure uniform conditions. The tests were given by thirty students from the New York Training School for Teachers who had taken a course in educational measurements and the normal course in muscular-movement penmanship. These girls were familiar with the various scales in penmanship and had had some experience in giving standardized tests. They also received special practice in giving this penmanship test, considerable attention being given to "timing."

The children memorized on the day before the test the sentence, "One must exercise in work and in play." The test consisted in writing this sentence repeatedly for two minutes. The class teachers took no part in the testing, although they were present at the time.

The test was given in fifteen schools involving five hundred classes and eighteen thousand children in grades IV to VIII.[3] The schools were chosen as representative of the whole system. About twelve thousand specimens were scored in full.

Training the judges.—The scoring of these twelve thousand samples of penmanship in each of the elements of form, movement,

[2] Lister, C. C. and Myers, G. C., "An analytic scale of handwriting," *Journal of Educational Psychology*, 9:417–31, October, 1918.

[3] In New York, A classes are low and B classes high divisions of the yearly grades.

spacing, and speed represented an enormous task. It was undertaken by Dr. Hamilton and Miss Scheuermann of the New York Training School for Teachers and by Mr. Lister of the Brooklyn Training School for Teachers. The services of training-school students were enlisted.

While a penmanship scale in the hands of trained scorers reduces variability in judgment, in the hands of untrained persons it gives no guarantee that judgments will be reliable. Nor does a knowledge of the theories underlying educational measurements or of the statistical derivation of scales insure accurate scoring. Likewise expertness in teaching penmanship while of some help does not imply reliability in scoring. To obtain accuracy in scoring there must be training in the use of the scale.

The training-school students who acted as judges had the following qualifications: (1) They had been instructed in the muscular-movement method. (2) They had taken the normal course in teaching that method. (3) They had taken the course in educational measurements at the training school.

It was arranged to give these students specific training for three weeks or longer through actual practice in judging specimens. For use in training the judges, standard specimens of writing were scored for form, movement, and spacing by one of the authors of the scale and by four of his associates. The specimens were coded so that the students were unaware of the "expert" score. In scoring these practice samples and subsequently in scoring the test papers the ascending-descending method was used.

It was planned to give the student judges an initial test in scoring and at the end of the practice period a final test. The data resulting from these tests would afford a measure of the reliability of each prospective judge and a basis for eliminating some of the students from participation in scoring the test papers. Unfortunately school conditions and other circumstances prevented us from carrying out our program in its entirety. We were, however, able to get an initial test for all the student judges and a final test for part of the group.

Table I shows for the initial test the distribution of the judges according to their average deviation per specimen from the expert scores. The table reads: In judging form one person deviated on the average 1 to 2 points from the expert scores, four

deviated 4 to 5 points, etc. From the last column it is evident that the greatest variation occurred in judging movement.

Details not shown in Table I indicate that scores above and below those of the experts were about equally frequent and of about the same range. Without regard to direction the amounts of the deviations at the time of the initial test were as given in Table II. Here as elsewhere a "point" is a unit of the scale. A step is the difference between successive samples on the printed scale and this difference is always ten points.

Under *Form* it will be seen that 878 or 78.8 percent of the judgments differed by one scale step or less from the standard or expert scores. In movement 69.6 percent of the judgments deviated one step or less, and in spacing 81.3 percent. In other words about three-quarters of the judgments either agreed with those of the experts or assigned values to form, movement, or spacing corresponding to the adjacent samples on the scale.

After the practice periods even this amount of unreliability was reduced. Complete data were secured for 56 of our 86 prospective judges. Their average deviation from the expert scores was reduced from 7.5 points, or three-quarters of a scale step, to 4.2 points, or less than half a scale step. Forty-two of them improved with practice, three remained stationary, and eleven lost ground—the last in each case but by a small amount.

Although the bulk of the papers were scored in each case by a single judge, an extensive study of the reliability of the ratings was made by having the papers of about 15 percent of the classes scored by three judges. A detailed analysis was made of the results for 20 classes, from which 753 specimens had been secured. It was found that on the average a judgment by one judge on an individual specimen varied less than four-tenths of a scale step from the median of three independent judgments. When individual results were combined in classes, it was found that the average error was less than three-tenths of a scale step. Since the object was not to secure results for individual pupils nor even for classes, but for larger grade groups of about 1,000 pupils in each case, it is believed that the trained judges rendered a reliable verdict for the intended purpose.

General results.—Table III shows the median scores obtained by each grade group throughout the city in each of the elements

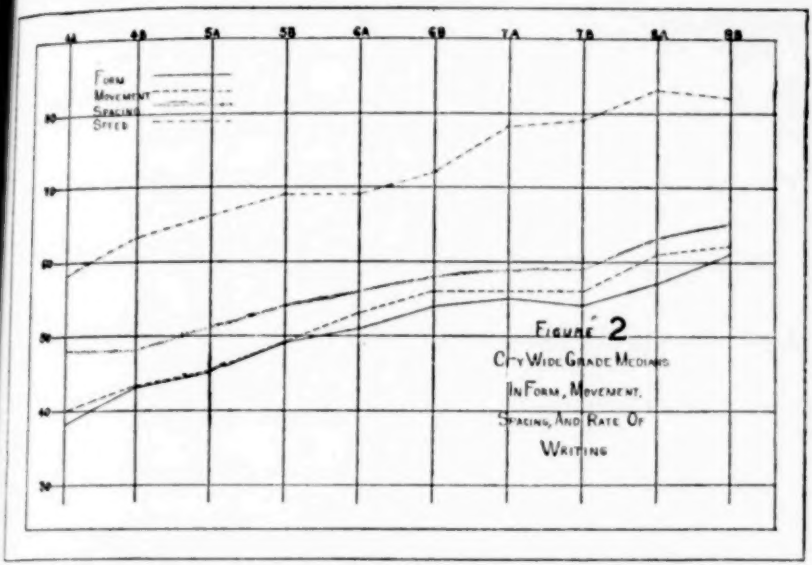
measured by the scale, in form, movement, spacing and rate of writing.

TABLE III. CITY–WIDE MEDIAN SCORES BY GRADES FOR FORM, MOVEMENT, SPACING, AND RATE

Grades	Form		Movement		Spacing		Rate		Number of Pupils
	Median	S. D.	Median	S. D.	Median	S. D.	Median	S. D.	
VIIIB	61	11.5	62	12.2	65	10.7	82	19.3	988
VIIIA	57	13.1	61	14.1	63	12.4	83	18.4	1,117
VIIB	54	12.2	56	12.7	59	11.1	79	19.2	1,225
VIIA	55	12.3	56	12.9	59	11.6	78	17.4	1,327
VIB	54	11.9	56	12.7	58	11.3	72	20.3	1,481
VIA	51	12.4	53	13.1	56	11.7	69	18.1	1,414
VB	49	12.6	49	13.2	54	12.1	69	17.6	1,327
VA	45	11.5	45	12.2	51	11.9	66	17.8	1,529
IVB	43	10.7	43	11.5	48	11.2	63	15.6	1,110
IVA	38	11.1	40	11.7	48	11.0	58	13.6	241
City	51.3	12.0	52.9	12.5	56.4	11.5	72.5	18.5	11,759

Figure 2 shows the medians of Table III graphically. Both the scale values and the rate in letters per minute are shown on the vertical axis. The horizontal axis shows the grades. For the city as a whole the median score for all the pupils tested was (without regard to grade) 51.3 in form, 52.9 in movement, and 56.4 in spacing. This means that 50 percent of the pupils tested, without regard to grade, obtained these scores or better. The table also shows that 50 percent of all the pupils wrote at a rate of 72.5 or more letters per minute.

The product of the school may be best judged by the work done in the highest grade. Figures 3, 4, and 5 indicate approximately the results obtained in the upper eighth grade. If we arranged all the samples from this grade in order from the best performance to the poorest, and if we counted down one-quarter of the way from the best toward the poorest we should reach a sample like Sample A of Figure 3. The point thus reached is called the upper (third) quartile. Similarly Sample B of Figure 4 represents the median or the point reached by counting down half way; and Sample C of Figure 5 represents the so-called lower



(first) quartile or the point reached by counting down three-quarters of the way.[4]

Table IV gives for each grade[5] the median rate, the upper and lower quartiles and the range of the middle 50 percent. It also gives, for comparative purposes, the standards set up by Ayres and by Freeman and the results obtained in the Cleveland survey. Figure 6 shows the facts graphically.

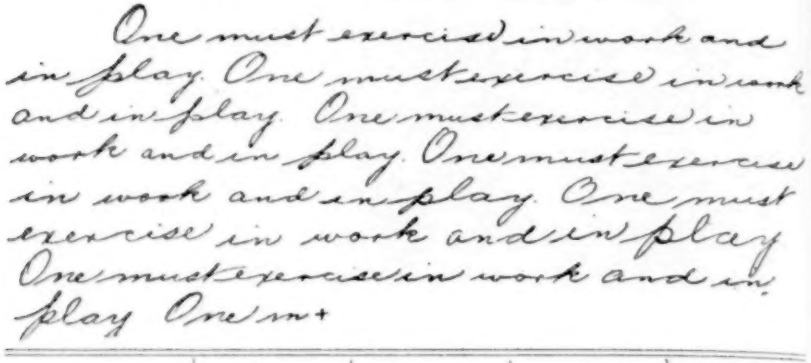
TABLE IV. GRADE MEDIAN RATES, UPPER AND LOWER QUARTILES

Grades	IVB	VA	VB	VIA	VIB	VIIA	VIIB	VIIIA	VIIIB
Q_3	73.0	77.3	79.3	79.8	86.7	91.8	93.7	95.4	94.4
M	63.3	66.3	69.1	68.7	72.2	77.8	79.4	82.5	81.5
Q_1	51.4	53.1	57.5	56.3	59.3	68.0	68.1	70.9	67.7
Range Middle 50 percent	21.6	24.2	21.8	23.5	27.4	23.8	25.6	24.5	26.7
Standards:									
Freeman	56	65		72		80		90	
Ayres	56	64		70		76		80	
Cleveland		60		70		76		80	

[4] The abbreviation for the lower or first quartile is Q_1; for the upper or third quartile, Q_3.

[5] The IVA data are omitted because of the few cases.

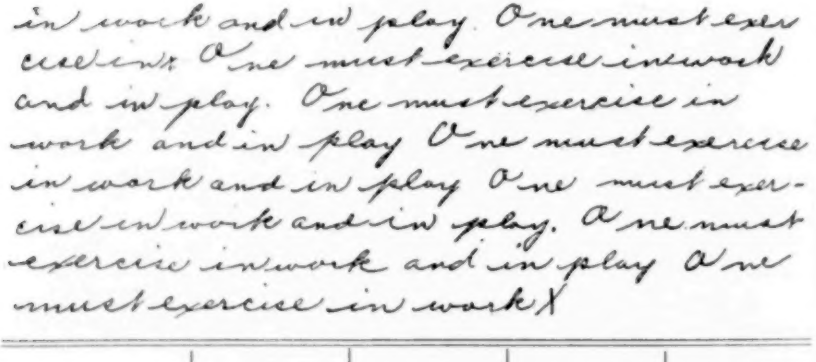
FIGURE 3. SAMPLE A, ILLUSTRATING THE TYPICAL PRODUCT OF THE BETTER HALF OF THE UPPER EIGHTH GRADE



	Form	Movement	Spacing	Speed
Sample A.....	70	70	70	92
Q_3.....	69	70	71	94

This sample was equaled or surpassed by 25 percent of the pupils of the upper eighth grade.

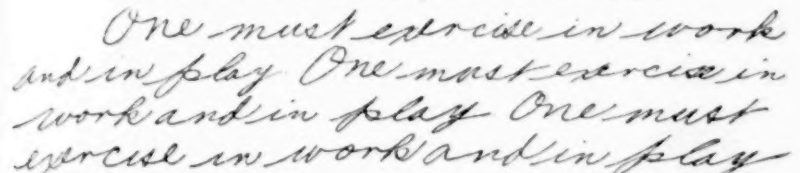
FIGURE 4. SAMPLE B, ILLUSTRATING THE TYPICAL PRODUCT OF THE UPPER EIGHTH GRADE



	Form	Movement	Spacing	Speed
Sample B.....	60	60	65	82
Median.....	61	62	63	82

This sample was equaled or surpassed by 50 percent of the VIIIB group.

FIGURE 5. SAMPLE C, ILLUSTRATING THE TYPICAL PRODUCT OF THE POORER HALF OF THE UPPER EIGHTH GRADE


One must exercise in work
and in play One must exercise in
work and in play One must
exercise in work and in play
One must exercise in work
and in play One must

	Form	Movement	Spacing	Speed
Sample C.....	55	50	55	64
Q_1.....	53	53	53	68

This sample was equaled or surpassed by 75 percent of the pupils of the upper eighth grade.

It appears that the median rate of writing increases irregularly from grade to grade, rising from 63 letters per minute in the IVB grade to 82 in the second half of the eighth grade. There is no gain shown in the VIA grade over the VB, while in the VIIIB the rate is slightly lower than that of the grade below. The greatest gain appears in the VIIA.

New York children are by no means slow writers. Their rates are higher than the scores obtained in other surveys and exceed in all years the standards set up by Ayres. With the exception of the last three half years, the median scores are higher than the standards fixed by Freeman which represent the average of the better half of the cities tested by him.

There is an astonishing variation within grades in respect to rate of writing. Even when classes rather than pupils are considered and median scores rather than individual scores, the range is still conspicuous. The typical performance of one class of the lower fifth grade was 32 words per minute, of another 88 words per minute. In the upper sixth grade one class scored but 39 while another scored 98. Even in the upper eighth grade the class scores ranged from 48 to 96.

FIGURE 6
MEDIANS - UPPER AND LOWER QUARTILES FOR EACH GRADE IN RATE OF WRITING

The condition is such as to indicate an absolute lack of grade standards. It probably shows that no particular attention has been paid to controlling the rate at which children write. Those who began as slow writers doubtless continued as slow writers throughout the grades, with only such increase as comes with maturity. The fast writers in the lower grades probably continue likewise to write at a rapid rate throughout.

While writing of a good quality is a prime requirement, instruction should also seek to improve the rate of writing, particularly

that of the slower writers. Less than 80 letters per minute is slow writing. The presence in the last year of the course of numbers of pupils who fail to attain a speed of 60 letters per minute indicates the failure of our instruction to give pupils about to leave school writing habits which will meet the requirements of the outside world. The presence of pupils in all grades who write at very rapid rates again indicates the need of control so that the quality of the writing of such pupils may not be sacrificed for the sake of speed. Rapid writing must be accompanied by satisfactory quality.

Quality of writing.—Since the New York scale records quality under three aspects—form, movement, and spacing—we shall devote a table to the most significant grade scores for each of these characteristics. Table V deals with *form*, Table VI with *movement*, and Table VII with *spacing*. Attention is called to certain facts which each of these tables exhibits:

1. The range from the median of the lower fourth grade to that of the upper eighth grade is small, amounting to about two steps on the scale or merely to the difference between adjacent samples in Figure 1.

2. The range of the middle 50 percent in each grade is almost as great as the range between the medians of the highest and lowest grades represented in the tables. In other words the difference between the performances of two fairly representative pupils of the same grade may be as great as the difference between the performances of two pupils one of whom is typical of the fourth grade and the other of the eighth.

3. Grades are scarcely distinguishable. This is shown not only by the small intervals between grade medians—intervals too small to be perceived on the scale—but also by the large degree of overlapping.

The extent of this overlapping may be roughly indicated by a comparison of the *upper* quartile of a grade with the medians of higher grades and by a similar comparison of the *lower* quartile of a grade with the medians of lower grades. These comparisons may be made directly from Tables V, VI, and VII. Table VIII shows for the element of *form* how the comparisons may be arranged.

In Table VIII the medians for the various grades are given across the top, the lower quartile is shown in the left-hand column, and the upper quartile in the last column on the right.

TABLE V. GRADE MEDIANS, UPPER AND LOWER QUARTILES IN FORM

Grade	IVA	IVB	VA	VB	VIA	VIB	VIIA	VIIB	VIIIA	VIIIB
Q_3	47.1	49.8	53.4	58.2	59.5	62.3	63.6	62.9	67.8	68.8
Median	37.9	43.2	45.4	49.2	50.5	53.7	54.5	54.0	57.0	60.8
Q_1	30.0	35.8	38.3	40.9	42.4	45.2	45.6	45.4	47.7	52.6
Range of middle 50%	17.1	14.0	15.1	17.3	17.1	17.1	18.0	17.5	20.1	16.2
Intervals between medians	5.3	2.2	3.8	1.3	3.2	0.8	0.5	3.0	3.8	

TABLE VI. GRADE MEDIANS, UPPER AND LOWER QUARTILES IN MOVEMENT

Grade	IVA	IVB	VA	VB	VIA	VIB	VIIA	VIIB	VIIIA	VIIIB
Q_3	48.9	53.1	54.6	59.7	63.3	65.3	65.5	64.7	69.9	69.6
Median	39.9	43.3	44.8	49.3	52.9	56.1	56.0	55.7	60.8	62.0
Q_1	33.3	35.7	36.4	39.8	43.1	46.3	45.9	46.4	49.2	52.9
Range of Middle 50%	15.6	17.4	18.2	19.9	20.2	19.0	19.6	18.3	20.7	16.7
Intervals between medians	+3.4	+1.5	+4.5	+3.6	+3.2	−0.1	−0.3	+5.1	+1.2	

TABLE VII. GRADE MEDIANS, UPPER AND LOWER QUARTILES FOR THE ELEMENT OF SPACING

Grade	IVA	IVB	VA	VB	VIA	VIB	VIIA	VIIB	VIIIA	VIIIB
Q_3	58.7	56.3	58.4	62.5	64.6	65.9	66.9	66.6	70.7	71.6
Median	47.7	47.7	50.5	54.0	56.2	57.7	58.7	58.7	63.2	64.8
Q_1	40.3	40.6	42.2	45.0	48.2	50.8	51.0	51.6	54.3	57.5
Range of middle 50%	18.4	15.7	16.2	17.5	16.4	15.1	15.9	15.0	16.4	14.1
Intervals between medians	0.0	+2.8	+3.5	+2.2	+1.5	+1.0	0.0	+4.5	+1.6	

TABLE VIII. COMPARISON OF QUARTILES WITH MEDIANS TO SHOW OVERLAPPING

Lower Quartiles		Medians										Upper Quartiles	
		IVA	IVB	VA	VB	VIA	VIB	VIIA	VIIB	VIIIA	VIIIB		
		37.9	43.2	45.4	49.2	50.5	53.7	54.5	54.0	57.0	60.8		
IVA	30.0		3.9	2.3								47.1	IVA
IVB	35.8	−2.1		4.4	0.6							49.8	IVB
VA	38.3		−4.9		4.2	2.9						53.4	VA
VB	40.9		−2.3	−3.6		7.7	4.5	3.7	4.2	1.2		58.2	VB
VIA	42.4		−0.8	−3.0	−6.8		5.8	5.0	5.5	2.5	1.3	59.5	VIA
VIB	45.2			−0.2	−4.0	−3.6		7.8	8.3	5.3	1.5	62.3	VIB
VIIA	45.6				−3.6	−4.9	−8.1		9.6	6.6	2.8	63.6	VIIA
VIIB	45.4				−3.8	−5.1	−8.3	−9.1		5.9	2.1	62.9	VIIB
VIIIA	47.7				−1.5	−2.8	−6.0	−6.8	−6.3		7.0	67.8	VIIIA
VIIIB	52.6						−1.1	−1.9	−1.4	−4.4		68.8	VIIIB

The figures to the right of the series of blank compartments which run diagonally across the table show how much the upper quartile of a given grade exceeds the median of the grades above. The figures to the left show how much the lower quartile of a given grade falls below the median of lower grades. For instance, reading horizontally across from VB we note that the lower 25 percent of the VB group fails to reach the median of the IVB grade by 2.3 points and the median of the VA grade by 3.6 points. The figures on the same line to the right of the diagonal of blank compartments show that the upper quartile of the VB exceeds the VIA median by 7.7 points, the VIB median by 4.5 points, the VIIA median by 3.7 points, the VIIB median by 4.2 points, and the VIIIA median (five grades above) by 1.2 points.

Relation of speed and quality.—Hitherto we have considered speed and quality separately; we shall now consider the relation between them. It is possible to set up grade norms for the various elements of quality, norms which may be achieved only at the expense of speed. We may also establish certain norms for speed which may likewise be attained only at the expense of quality. Satisfactory quality must be attained, but it must be accompanied by satisfactory rates of writing.

From Table III which showed the medians in each characteristic of quality and also in rate, it may be noted that speed increased with one or two exceptions from grade to grade. At the same time the median scores in form, in movement, and in spacing also increased from grade to grade. There was thus a certain degree of correspondence as far as the whole grade groups were concerned, between the median speed and the median quality from grade to grade.

Does such relation hold within the grades? In order to determine this the average speed was determined for each scale score in each grade in form, movement, and spacing. Table IX shows for each grade the average speed at which each quality in form was written. This table reads: In the IVB grade 91 pupils scored 20 in form and their average rate of writing was 63 letters per minute, etc.

For each grade there was a general tendency toward decrease of speed as *form* improved. From data not here reproduced similar facts were discovered with reference to *movement* and *spacing*.

TABLE IX. AVERAGE RATE IN EACH GRADE FOR EACH STEP ON SCALE IN FORM

Grade	Score in Form															
	20		30		40		50		60		70		80		90	
	No.	Average Rate	No.	Average Rate	No.	Average Rate	No.	Average Rate	No.	Average Rate	No.	Average Rate	No.	Average Rate	No.	Average Rate
IVB	91	63	321	64	431	63	195	62	52	59	17	49	3	48	.....	.....
VA	94	73	346	71	599	66	313	62	124	59	45	52	8	45	3	48
VB	58	76	235	72	404	72	362	67	169	63	92	60	7	61	.....	.....
VIA	35	75	211	71	440	72	394	69	213	65	106	63	15	58	.....	.....
VIB	23	82	140	77	397	75	482	73	295	68	126	62	18	56	.....	.....
VIIA	25	87	119	85	336	82	408	80	297	76	121	73	20	73	1	45
VIIB	26	104	103	90	325	85	398	79	234	76	122	75	17	75	.....	.....
VIIIA	8	95	87	90	238	89	320	83	237	82	193	75	31	77	3	65
VIIIB	2	115	41	83	126	86	300	82	310	79	178	77	31	77	.....	.....

But although both rate and quality improved from grade to grade, we have no satisfactory answer to the general question of the relation between rate and quality of handwriting among school children. The correlation coefficient is the most exact expression of the relationship. Accordingly it was computed (by the product-moment formula) for each pair of characteristics. Instead of treating all the data involved in the scores of the 12,000 pupils, two sets of 1,000 specimens each were selected by a process of sampling. The results are shown in Table X.

TABLE X. COEFFICIENTS OF CORRELATIONS BETWEEN THE SEVERAL ELEMENTS

Elements	First Sampling		Second Sampling	
	r	P. E.	*r*	P. E.
Form and movement.....	+.75	.0093	+.749	.009[2]
Form and spacing.....	+.73	.0094	+.738	.0097
Movement and spacing....	+.72	.0101	+.716	.0103
Speed and form.....	−.056	.02	−0.54	.0[2]
Speed and movement.....	−.036	.02	−.029	.0[2]
Speed and spacing.....	−.013	.0213	−.016	.0213

The correlations between speed and the elements of form, movement, and spacing are negligible. The fact that the probable error is as large as or larger than *r* indicates the unreliability of our result. It appears then that there is little relation between speed and the different elements of quality in writing. The fast writers write a good hand as well as the slow writers. The good writers write rapidly as may also the poor writers.

In the case of the other elements a high degree of positive correlation is shown. This indicates that an increase in form would be accompanied by an increase respectively in movement and in spacing, and vice versa.

Quality and speed in different schools.—The results already presented have clearly indicated the variability of achievement in the various classes and grades. It has been shown that there is apparently an absolute lack of uniform standards. In this section data will be presented showing the relative emphasis placed upon speed and upon quality in some of the schools tested.

In order to show the development in penmanship through the grades, the progress in quality and the progress in speed were shown in one diagram. For quality the grade-median scores in form, movement, and spacing were averaged to obtain a single score. The resulting curves would not be materially different if we were to take only one element at a time.

Figure 7 shows the progress from grade to grade in three selected schools. The horizontal axis shows speed, and the vertical axis quality. In interpreting the diagrams we must bear in mind that the distance between the adjacent horizontals measures only ten points or one step on the penmanship scale, and that distance between the verticals represents but ten letters.

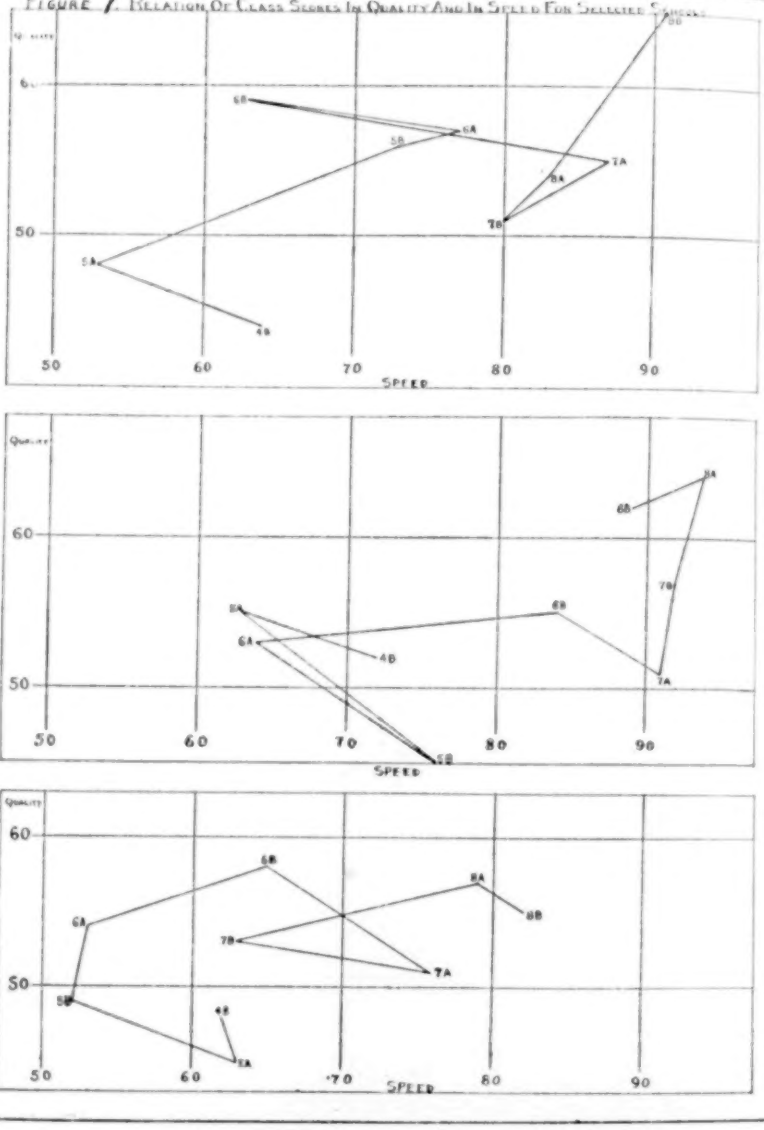
The reader would do well to trace the curious paths from IVB to VIIIB as shown in Figure 7. For example at the school represented by the top section of the figure we start from IVB to VA by going backwards in speed, while making some (but not much) headway in quality. From VA to VB the lost ground in speed is more than recovered and there is a further gain in quality; VIA adds little in either quality or speed; VIB shows a loss in speed with very little gain in quality; etc. From such representations as these—and space alone limits our presentation of others—we get the impression of haphazard results. The condition is what might be expected in the absence of standards and of means of control.

Tentative Grade Standards

The aim of penmanship in our elementary schools is to give the pupil a certain degree of writing skill. The course of study at New York prescribes that seventy-five minutes a week be devoted to penmanship in the IA to VIB grades, and that in that time the child be trained to write with legibility, ease, and speed. Such aims, however, are vague and indefinite and offer no material guidance to the teacher or principal in his work in penmanship.

Prior to the derivation of scales for objective measurement, our aims in penmanship, as in other subjects, were necessarily expressed in general statements rather than in definite terms. Now that we have such scales and means of measurement, we can formulate exactly the degree of writing skill that a pupil should attain as a result of the elementary course.

FIGURE 7. RELATION OF CLASS SCORES IN QUALITY AND IN SPEED FOR SELECTED SCHOOLS



The end product of our school training is represented by the VIIIB pupil who is about to graduate. How well should a pupil in the VIIIB write? At what speed? Immediately we answer such questions by asking other questions. How well need he write? How fast need he write? We find our answers in the demands of society upon the pupil and in the needs of his later life.

A number of investigations have been conducted to determine the degree of writing skill required by the outside world in various lines of endeavor. Such studies, however, were reported in terms of the Ayres scale so that the results could not be used in setting up norms for the New York scale. An independent attempt was made with the New York scale, to obtain under test conditions samples of handwriting of clerical employees in several large department stores. The group tested numbered 161, distributed as follows: 98 salesclerks, 17 billers, 39 checkers, and 7 clerks and bookkeepers. In all these positions legible, rapid writing was considered essential.

Table XI presents the median scores in each element for each of the above groups in comparison with the median scores of our VIIIB pupils.

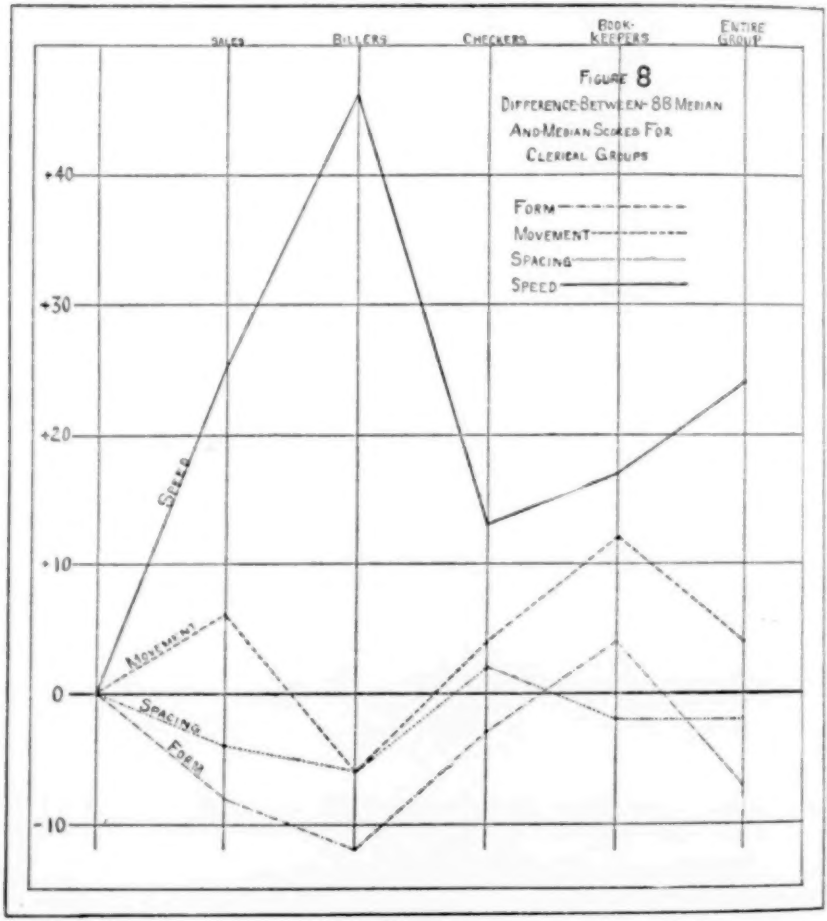
TABLE XI. MEDIAN SCORES OF GROUPS OF CLERICAL EMPLOYEES IN DEPARTMENT STORES

Group	Number	Form		Movement		Spacing		Speed	
		Median	Difference from VIIIB	Median	Difference from VIIIB	Median	Difference from VIIIB	Median	Difference from VIIIB
VIIIB Pupils	988	61	.....	62	.....	65	.....	82	.....
Salesclerks.....	98	53	− 8	68	+ 6	61	−4	107	+25
Billers.....	17	49	−12	56	− 6	59	−6	128	+46
Checkers.....	39	58	− 3	66	+ 4	67	+2	95	+13
Bookkeepers....	7	65	+ 4	74	+12	63	−2	99	+17
Entire outside group.....	161	54	− 7	66	+ 4	63	−2	106	+24

The data in Table XI are presented in Figure 8 in terms of the amounts by which each employee group exceed or fall below the

VIIIB medians. It will be noted that in the three elements of quality the VIIIB pupils did quite as well as the salespeople. In speed, however, the employee group greatly exceed the rate of the VIIIB pupils, the salespeople writing at the rate of 107 letters per minute or 25 letters above the VIIIB median, the billers 128 letters or 46 above, the checkers 95 or 13 above, and the bookkeepers 99 or 17 letters above the VIIIB rate.

The data presented above while suggestive are not regarded as conclusive, and it is planned to continue the investigation. Meanwhile, Table XII presents the grade standards proposed for



each element. They are shown in Figures 9 and 10 in comparison with the upper quartiles and city-wide medians.

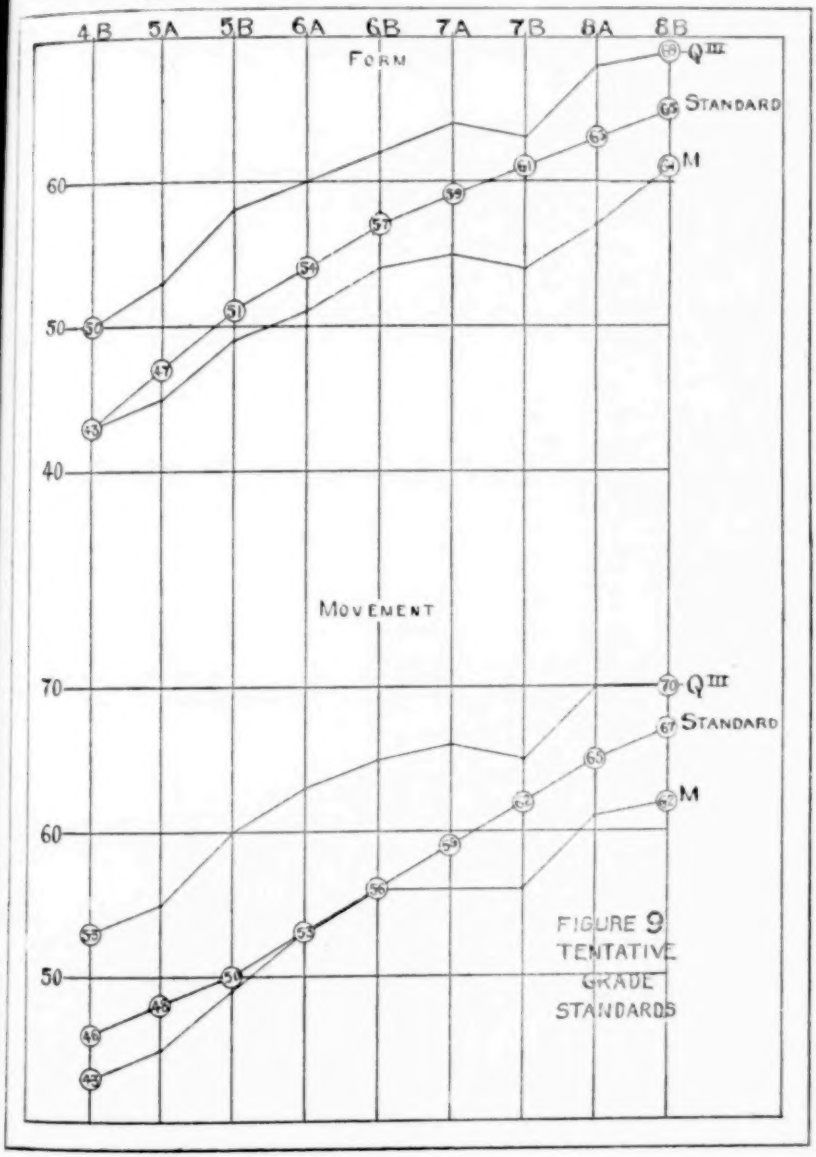


FIGURE 9
TENTATIVE
GRADE
STANDARDS

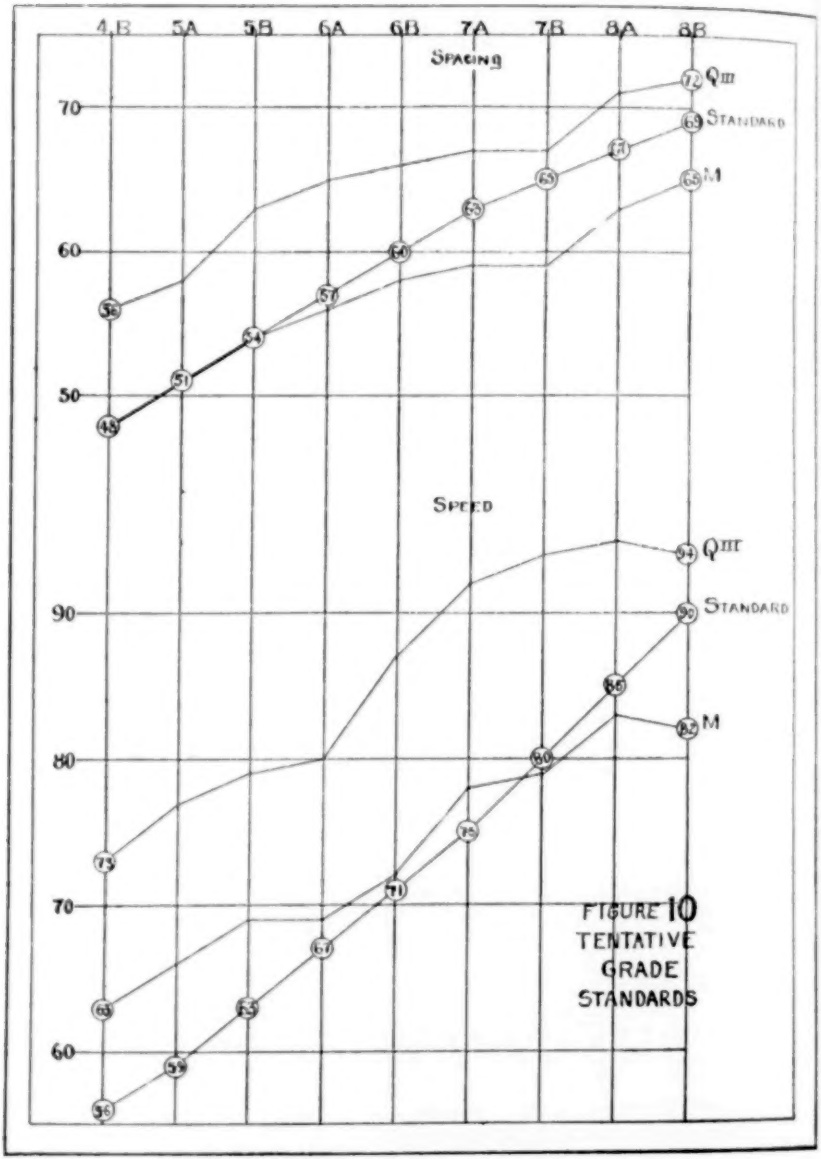


FIGURE 10
TENTATIVE
GRADE
STANDARDS

TABLE XII. TENTATIVE STANDARDS IN PENMANSHIP FOR THE NEW YORK PENMANSHIP SCALE

Grade	Form	Movement	Spacing	Speed
IVB	43	46	48	56
VA	47	48	51	59
VB	51	50	54	63
VIA	54	53	57	67
VIB	57	56	60	71
VIIA	59	59	63	75
VIIB	61	62	65	80
VIIIA	63	65	67	85
VIIIB	65	67	69	90

For the VIIIB standard in form the average of the scores attained by the classes which exceeded the city-wide grade median was taken. This score of 65 is already reached by 35 percent of the VIIIB groups tested or by 10 of the 33 classes involved. The lower end of the curve was begun at 43, the city-wide median for the IVB. The curve of the grade standards was raised from the median curve, in some places arbitrarily, and in other places through smoothing to the points indicated. In the VIB, VIIA, and VIIB, the grade medians were on a level, whereas the "Standard" curve maintains an even progress. The same procedure was followed in the elements of movement and spacing.

For speed the standards adopted are taken from those set up by Freeman. Figure 10 shows that in the lower grades such speed norms are below the grade medians.

The above standards are frankly tentative; but when the achievements of the better schools are considered they are not regarded as unreasonable nor difficult of accomplishment. The schools are working under a uniform time allotment and what is accomplished in some of the schools ought to be attainable elsewhere.

A COMPARATIVE STUDY OF THE INTELLIGENCE OF WHITE AND COLORED CHILDREN

R. A. SCHWEGLER
School of Education, University of Kansas
AND
EDITH WINN
High School, Neodesha, Kansas

Recent events in our national history have once more forced the question of the education of the American negro sharply into the foreground. Much has, of course, already been done by way of providing educational facilities for negro children. In the north there has been little or no attempt at differentiation. The white and the colored child have been freely admitted to equal educational opportunity. The south naturally has viewed its problem from a different angle, but even there state agencies have been awakened by patriotic, religious, and philanthropic forces to a new and productive interest in education for the colored race, until it may be said that at least the administrative aspects of the matter are in a fair way of being solved.

But the experience of unbiased educators has uniformly been that for some reason the negro child does not function in the American school as the white child does. All available statistics point to a higher school mortality, and to a markedly higher ratio of retardation. Various reasons have been given for this state of affairs. It has been asserted that the social status and traditions of the negro encourage meager intellectual standards, and that this is reflected in the school achievements of the negro child. By others it has been claimed that race differences of a fundamental type are involved, that the mentality of the negro child is essentially different from that of the white child, making identical standards both unfair and unwise. Of all the many explanations offered, these two have been most persistently and plausibly stated. It seems wise to attack the problem by a series of careful, unbiased investigations. Prejudiced opinion is rarely a good vehicle for truth.

This paper is an attempt to study the relative intelligence of white and colored children by the use of scientifically reliable

devices. It is perhaps only fair to say that the study was undertaken with a distinct bias in favor of the colored child. Three questions were raised: (1) what is the relative intelligence of the two races; (2) what is the relative intelligence of colored boys and girls; and (3) in what particular types of function, if any, is the difference between the two races most striking.

Among the first attempts to measure the mental endowment of the negro by an intelligence scale was that of Miss Alice C. Strong in 1913. The Binet test was given to 225 white and to 125 colored children of Columbia, South Carolina. Josiah Morse, in reporting the study, calls attention to the fact that while 29.4 percent of the colored children were found to be backward more than one year, only 10.2 percent of the white children were so retarded. He also points out that only 0.8 percent of the colored children were one or more years advanced, while 5.3 percent of the white children were thus advanced. Morse emphasizes the fact that the mental level of the colored children approximates that of a typical mill population. Discussing the question of special abilities, he calls attention to the fact that colored children excel in rote memory, naming words, making rhymes, and in time orientation, while they are clearly inferior to white children in aesthetic judgment, observation, reasoning, motor control, logical memory, use of words, resistance to suggestion, and in orientation to the complexities of civilized society. He finds also that according to the Binet scale, a somewhat larger number of white children are in a grade below than in a grade above the one corresponding to their mental ability, and that the reverse is true of colored children. Morse concludes his remarks on Miss Strong's study with the statement that "negro children from 6 to 12 and possibly to 15 years are mentally different from, and also younger than, southern white children of corresponding ages."

B. A. Phillips reports the results of a comparative study of negro and white children in the elementary schools of Philadelphia. Both mental and environmental factors were here taken into account, and it was found that whereas 42.1 percent of white children were retarded, 68.4 percent of the colored children were so handicapped. A general survey of the tests again shows that "colored pupils as a class were good in memory tests and poor in those requiring judgment."

Lacy, in a study of one hundred retarded pupils in the fourth grade at Evanston, Illinois, found that the differences in the distribution of the intelligence quotients of the colored and white children was very marked, and that the median intelligence quotient of the white children was at every age superior to that of the colored children. "On the whole the intelligence of the colored children is inferior to that of the white children."

A variety of tests other than the Binet scale has been employed in studying the relative intelligence of the two types of children. Thus, studies employing tests of cancellation, quickness of learning, substitution, completion, association, and memory are reported. The findings are all substantially as above reported. For fuller details the reader is referred to the bibliography appended to this paper.

The results herewith reported are derived from a careful study of 116 pupils drawn from the seventh and eighth grades of the junior high school of Lawrence, Kansas. The total colored enrollment in this school was 58, there being 24 boys and 34 girls. A chance selection of an equal number of white boys and girls was made from the enrollment cards. The names of the children so selected were, with one exception, wholly unknown to the examiners. These two groups of children were tested under uniformly favorable conditions by the use of the Stanford-Binet scale, arranged for use as a point scale. Each child was given a chance at every test, in order to assure perfect fairness, and to secure all the advantage of wide range testing.

The children were tested singly, the time required for each child varying from one to two hours, depending upon the promptness of the child's response to each test. The work was done during school hours in the school building during the first four months of 1919. It is of interest to note that on the whole the colored children appeared more at ease with the examiner than did the white children. Some of the white children seemed nervous and fearful lest they fail in the test. *The fear of failure seemed to be alien to the colored children's minds.* For the most part, they manifested an attitude of willingness to cooperate, and they seemed quite as interested and attentive as were the white children.

Considering the fact that at Lawrence the two races have identical school facilities throughout, and that environmental

conditions are perhaps as nearly equal as they are anywhere in the United States, the conclusions derived from this study should have at least a fair degree of reliability.

The distribution of chronological ages of the children selected for this study may readily be seen from Table I.

TABLE I. DISTRIBUTION OF CHRONOLOGICAL AGES

Age	Colored		White	
	Girls	Boys	Girls	Boys
10.0–10.4	1			
10.5–10.9				
11.0–11.4				
11.5–11.9			4	3
12.0–12.4	2	1	5	1
12.5–12.9	5		6	2
13.0–13.4	4	5	5	1
13.5–13.9	6	3	5	4
14.0–14.4	4	4	5	2
14.5–14.9	4	4	2	4
15.0–15.4	2	3	2	2
15.5–15.9	3	1		2
16.0–16.4	2	2		2
16.5–16.9		1		
17.0–17.4				1
17.5–17.9	1			
Total	34	24	34	24
Median Age	13.92	14.38	13.20	14.25

A brief study of Table I reveals the fact that the distribution of the chronological ages of the white and colored children is fairly similar. The median age for the colored girls and boys is respectively 13.92 and 14.38, while that of the white girls and boys is respectively 13.20 and 14.25 years. The girls in each racial group are somewhat younger than the boys, though the difference between the median ages of the colored boys and girls is not as great as it is in the case of the white boys and girls. The median age of the colored group is 14.13, while that of the white group is 13.61.

The results of the mental tests as applied to these children are set forth in Table II, which shows the distribution of mental ages.

TABLE II. DISTRIBUTION OF MENTAL AGES

Mental Age	Colored		White	
	Girls	Boys	Girls	Boys
9.0– 9.4	1			
9.5– 9.9		1		1
10.0–10.4	2	3	2	1
10.5–10.9	2			
11.0–11.4	4	2	1	1
11.5–11.9	6	3		1
12.0–12.4	7	2	4	2
12.5–12.9	2	3	2	1
13.0–13.4	3	4	5	3
13.5–13.9	1		4	5
14.0–14.4	5	3	5	4
14.5–14.9	1	2	3	2
15.0–15.4		1	4	1
15.5–15.9			1	1
16.0–16.4			2	1
16.5–16.9			1	
Total	34	24	34	24
Median Age	12.14	12.67	13.88	13.70

The median mental age of colored girls according to this table is 12.14, that of white girls is 13.88. In like manner the median mental age of colored boys is 12.67 against 13.7 for white boys. The ratio between these median mental ages and the median chronological ages already presented in Table I can be tersely expressed in the form of median group intelligence quotients. For the colored girls, colored boys, white girls, and white boys these ratios are respectively, 0.872, 0.881, 1.052, and 0.961, from which it would appear clearly that there is an appreciable difference between the mental performance of the two racial groups.

The question can, however, be studied from another angle. Table III presents the distribution of the actual intelligence quotients derived from the test results. A study of this table should be useful, for it will bring to the surface in sharp relief the general tendencies which intelligence quotients seem to follow.

TABLE III. DISTRIBUTION OF INTELLIGENCE QUOTIENTS

Intelligence Quotient	Colored		White	
	Girls	Boys	Girls	Boys
60– 64	1			2
65– 69	1	4	1	1
70– 74	3		1	1
75– 79	4	1	1	1
80– 84	2	2	1	
85– 89	7	6	2	4
90– 94	7	7	2	1
95– 99	3		3	2
100–104	3	1	5	4
105–109	1	2	5	3
110–114	1		3	
115–119			2	
120–124		1	3	4
125–129			4	1
130–134				
135–139	1		1	
Total	34	24	34	24
Median I. Q.	89.3	89.2	106.0	100.0

The intelligence quotient may be defined as the ratio of mental age to chronological age. Since it is being proposed and rather widely accepted as a reliable "measure of brightness," Table III takes on added significance, for it appears that while the actual median quotient of white girls is 106, that of colored girls is only 89.3; and that while the median quotient of white boys is 100, that of colored boys is 89.2. If the groups be studied as wholes, then the intelligence quotient of the white children as a whole is 103.3, and that of the colored children as a group 89.2. Of the white children 74 percent exceeded the median I.Q. of the colored children, while only 12 percent of the colored children exceeded the median I.Q. of the white children.

Following Terman's grouping of intelligence quotients, Table IV may be constructed. This table is still more significant when it is understood that while the colored group represented *all* the children of that race in the Junior High School, the chance-selected white group "happened" to include a number of the dullest and none of the most brilliant. Some of the white children who were not included in the investigation had earned intelligence quotients well above 140.

TABLE IV. GROUPING OF COLORED AND WHITE CHILDREN ON THE BASIS OF BRIGHTNESS

Intelligence Quotient	Degree of Brightness	Colored	White
140 and above	Genius or near-genius	0	0
120–139	Very superior	2	13
110–119	Superior	1	5
90–109	Average	24	25
80–89	Dull	17	7
70–79	Borderline	8	4
Below 70	Defective	6	4

Of the colored children 53.4 percent fall below 90, the point designated by Terman as representing the lower limit of normal white intelligence. Of the white children only 26 percent fail to reach this level, which is quite compatible with a normal chance distribution. Of the white children 31 percent present superior intelligence or better, while only 5 percent of the colored children reach this level.

A careful analysis of the successes and failures in the individual tests of the Binet scale is illuminating, and may ultimately serve to throw some light on the problems raised in connection with the study of negro mentality. The facts are briefly as follows:

1. In all rote memory tests not involving reversing or other manipulation of reproduced material the two races were for the most part on an even footing, or in any case the difference was slight, and possibly due to chance factors.

2. In all memory tests that involved reversing or other manipulation of reproduced material the white children were markedly superior. The reason for this difference is difficult to see, unless there is involved an essential difference in function trends.

3. In all tests involving common sense adjustment to practical situations of a fairly familiar type the colored children were clearly the equal of the white children. In fact, in some of the tests the colored boys outdid the white boys by a good margin. For example, the problem of "the ball lost in the field" was solved in a superior way by 62.5 percent of the colored boys, and by only 41.66 percent of the white boys. The white girls, however, surpassed the colored girls 64.7 to 52.94.

4. In all tests involving adjustment to practical situations that were unique, or that involved the consideration of somewhat remote contingencies of no matter how simple a type, the white children of both sexes were definitely superior. The white children were superior in problems which Terman calls "Problems of the third degree," 100 to 94.82; in arranging five weights in order 86.2 to 67.24; in reconstructing dissected sentences 84.48 to 44.82.

5. All tests which involved abstract reasoning whether of analytic or synthetic type were passed more satisfactorily by the white children than by the colored children. Sometimes the margin was very wide, as in defining abstract words, 84.48 to 48.27, in discovering absurdities 91.38 to 69, and in noting differences between abstract terms, 41.38 to 12.

6. In all tests involving primarily verbal facility, the two races are practically on a par. There is a slight advantage on the side of the white children, but it is negligible.

7. In tests involving recognition of ideas, white children show a fairly clear advantage over the colored children. Terman's vocabulary test gives a mean of 38.48 words per colored child, and 48.04 per white child. If Terman's contention that the vocabulary test actually possesses diagnostic worth is ultimately established, then these performances take on a very definite meaning.

Table V presents in more concise form the specific findings in connection with the individual test groups above mentioned. From the values there stated, it becomes clear that each group tends to present characteristic capacity endowments, which may have a bearing on the economic, industrial, and social history of the two races.

It appears clear that in some types of mental function the colored group is quite the equal of the white group; but that in other functions, particularly in those which involve continuous and somewhat intense concentration of attention on subjectively developed and manipulated imagery the white group is clearly superior.

Incidental to the detailed study of the test results, it also appeared that while there was a clearly marked superiority of the white girls as compared with the white boys in nearly all the

TABLE V. THE RELATIVE EFFICIENCY OF COLORED AND WHITE CHILDREN IN MENTAL TESTS

Type of Tests	Average of Colored Passing	Average of White Passing	Ratio of Colored to White
Direct, non-manipulated reproduction (memory)	74.94	74.39	1.007
Reversed or otherwise manipulated reproduction	47.28	57.52	0.822
Familiar common sense adjustments (orientation)	93.61	94.84	0.988
Adjustment to unfamiliar situations	36.2	58.18	0.622
Abstract reasoning	43.09	72.41	0.595
Verbal facility (common sense type)	91.38	93.1	0.981
Vocabulary recognition	38.48	48.04	0.801

Total net efficiency of colored group as compared with white group, 83.1 percent.

tests of the series, there was no such clear superiority of the colored girls over the colored boys. Table VI, derived from the undistributed data, shows the situation clearly in terms of intelligence quotients.

TABLE VI. RELATIVE INTELLIGENCE OF COLORED BOYS AND GIRLS AND OF WHITE BOYS AND GIRLS

Group	Median I. Q.	Range	Variability
Colored boys	89.2	0.65 to 1.23	1.43
Colored girls	89.3	0.64 to 1.36	1.59
White boys	99	0.62 to 1.25	1.82
White girls	106	0.67 to 1.37	1.52

The physical and mental precociousness of the white girl undoubtedly accounts in part for the relation above revealed. The intelligence of the white boys in this group is 94 percent of that of the white girls. The intelligence of the colored boys is 100 percent of that of the colored girls. *One is tempted to ask whether perhaps the growth rhythm of the two races may not be essentially different,* and whether the well-recognized difference between the intellectual life of adolescent white boys and girls is also characteristic of colored adolescent children. The groups studied were too small to draw final conclusions, but a significant question is here raised.

The findings of this study coincide in all essential points with those published by others. There appears to be an unmistakable difference in the intellectual life of the two groups studied. The median intellectual endowment of the colored group is about 85 percent of that of the white group. This difference, however, does not extend over the whole range of mental performances, but is highly specialized. In some types of function, namely, those that involve direct unmanipulated recall, in common sense adjustments to simple situations, and in verbal facility the two groups are equal in ability. But in all functions that involve intensive concentration of attention on complex subjective activities of a rational type the white group has a distinctly greater ability. The difference lies essentially in the higher and more abstract types of psychic work. That the difference is fundamental is indicated by the very fact that the colored children approached the test cheerfully and without concern as to the outcome, while the white children were obviously concerned lest they "fail the test." One group functioned in terms of the obvious, the other in terms of the related possibilities.

This study labors under the disadvantage of serious limitations. The groups studied were small, and therefore subject to distortion. But most serious of all is the fact that the tests employed were only adequate for the study of intelligence and touched the large field of other psychic functions inferentially or not at all. Perhaps the colored group compensates in some particular for its shortcoming in intellectual function.

Bibliography

Baldwin, B. T. "The learning of delinquent adolescent girls as shown by a substitution test," *Journal of Educational Psychology*, 4:317–32, June, 1913.

Bruner, F. G. "Racial differences," *Psychological Bulletin*, 11:384–86, October 15, 1914.

Calloway, Clinton J. "Co-operative school building," *Southern Workman*, 48: 236–41, May, 1919.

Eggleston, Edward. *The ultimate solution of the American negro problem.* Boston: Richard G. Badger, 1913. 285 pp.

Ferguson, George O. *The psychology of the negro: an experimental study.* (Archives of Psychology, no. 36.) New York: Science Press, 1916, 138 pp.

Lacy, W. I. "A study of one hundred retarded fourth-grade pupils tested by the Binet scale," *Psychological Clinic*, 12:16-24, March 15, 1918.

LIND, JOHN E. *Diagnostic pitfalls in the mental examinations of negroes.* (Reprinted from the *New York Medical Journal*, June 27, 1914.)

MAYO, M. J. *The mental capacity of the American negro.* (Archives of Psychology, no. 28.) New York: Science Press, 1913. 70 pp.

MORSE, JOSIAH. "A comparison of white and colored children measured by the Binet scale of intelligence," *Popular Science Monthly*, 84:75–79, January, 1914.

Negro education. (Reprinted from United States Bureau of Education Bulletin, 1916, nos. 38–39.)

PHILLIPS, B. A. "Retardation in the elementary schools of Philadelphia," *Psychological Clinic*, 6:79–90, May 15, 1912.

PHILLIPS, B. A. "The Binet tests applied to colored children," *Psychological Clinic*, 8:190–96, December 15, 1914.

PYLE, W. H. "The mind of the negro child," *School and Society*, 1:357–60, March 6, 1915.

STETSON, G. R. "Some memory tests of whites and blacks," *Psychological Review*, 4:285–89, May, 1897.

STRONG, A. C. "Three hundred fifty white and colored children measured by the Binet Simon measuring scale of intelligence," *Pedagogical Seminary*, 20:485–515, December, 1913.

TERMAN, LEWIS M. *The measurement of intelligence.* Boston: Houghton Mifflin Co., 1916, 362 pp.

WASHINGTON, BOOKER T. *Working with the hands.* New York: Doubleday Page and Company, 1904. 246 pp.

EDUCATIONAL MEASUREMENT IN THE FIELD OF HISTORY

CHAS. L. HARLAN
State Normal School, Lewiston, Idaho

Those who advocate the use of educational tests and scales must look upon the product, outcome, or attainment being tested as reduced to its lowest terms. Spelling attainment is tested by discovering whether the letters of a given set of words are in the correct order. Handwriting consists of letter forms of a given slant, spacing, heaviness of line, and alignment. Composition consists of words, sentences and paragraphs, spelled, capitalized, punctuated and arranged so as to express ideas more or less clearly, grammatically, and impressively. A similar situation prevails when one analyzes the other subjects for testing purposes. The tests and scales which have been most useful to the teacher and supervisor in diagnosing educational situations have been tests of specific attainment, those which indicate definitely the particular phases of a given subject in which a pupil is succeeding or failing. If the teacher, after testing her pupils in language, arithmetic, or spelling is not better informed concerning the specific weaknesses and strengths of her pupils, if she has not learned what particular phases of language, what processes of arithmetic, or what words in the spelling work need more attention in her teaching, the work of testing has not been greatly profitable.

To one seeking to devise standard tests in history, geography, literature and the other so-called "content" subjects, unusual difficulties are presented. The field of history with its many principles, facts, and details is almost unlimited in its extent. It is difficult to determine what the minimum essentials are. No teacher expects her pupils to assimilate and retain all the details of any history lesson. Yet there is little agreement as to what facts and principles a pupil should permanently acquire. There are some even who would claim that they do not teach history at all for the facts and principles which it contains but rather for the ideals and points of view which it may engender. Moreover, the content of history is being constantly and indefinitely increased by the addition of new material. It is difficult, for example, to foretell how much the accepted content of American history will be extended by the material dealing with the World War.

Another difficulty is apparent to the maker of standard tests in the field of history, namely, the failure of writers and teachers of history to define in commonly accepted and definite terms what the outcomes of the study of history should be. Information concerning the great men, the important movements, the dominant institutions of the past is interesting and inspiring; and this fact seems to one group of history teachers to justify the presence of history in the course of study. That the subject furnishes an excellent means for developing ideals of patriotism, heroism, courage, and honesty is sufficient justification to another group of history teachers. A third group argues that a pupil by viewing in perspective the growth and development of our institutions may thus secure a rational basis for his own actions as a citizen. These and other conflicting views as to why we teach history lead to indefiniteness and uncertainty as to the desirable outcomes of teaching, and consequently difficulty in determining the minimum essentials in this field.

Bagley and Rugg[1] have aided materially in defining the scope of the teaching of history. In most schools the textbook is the chief source of information in history. Certain names, places, events, dates, periods, movements, principles, and terms are presented in all the texts, while certain others occur with high degrees of frequency. These constitute a common basis in content which is at least presented to any reader of any of those commonly used texts. These common elements may therefore be considered as the minimum for testing purposes. Moreover, no matter what the aims of history teachers may be, one can be sure that this minimum of *information* has been given to the pupils.

Successful achievement in any subject implies at least the acquisition, by the pupil, of the essential knowledge in that subject. There may be many other outcomes from the study of history such as ideals, perspectives, points of view, and appreciations, but in the last analysis they are all dependent upon the essential facts and body of information presented to the pupil.[2] Therefore the incorporation of a part of this body of information is a prime requisite in the making of a standard history test.

[1] Bagley, W. C. and Rugg, H. O. *The content of American history as taught in the seventh and eighth grades.* (University of Illinois Bulletin, v. 13, no. 51, August 21, 1916.)

[2] Buckingham, B. R. "Proposed index of efficiency in teaching United States history," *Journal of Educational Research*, 1:161–71, March, 1920.

The multiplicity of aims in teaching history has led to a variety of methods of presentation. The earlier methods were designed to appeal to memory. Later ones furnish problems for thought and reasoning. Others seek to have the pupil organize and judge of the values of the information presented. Still others seek nothing more than to have pupils appreciate the heroism, the courage, and the sacrifices of the makers of history.

In order to meet this variety of methods and points of emphasis a history test should call for the exercises of a variety of mental functions and processes. Exercises which test memory alone are not sufficient. In addition to tests of memory there should be tests of association, organization, generalization, inference, judgment, comprehension of terms. This calls for variety in the types of exercises used in testing.

Another condition to be met with in making a satisfactory history test is that of eliminating, in so far as possible, elements of other subjects such as reading, composition, and spelling. A test of achievement in any of the "content" subjects should, as much as possible, be independent of attainments in the other subjects. A history test which requires much silent reading before the exercises can be worked out is a test of achievement not only in history but also in silent reading. Likewise, a test which requires long, involved or elaborate answers is, in part at least, dependent upon the pupil's composition ability. A pupil may be only a mediocre silent reader and a very poor writer of compositions; yet he may be rather proficient in his reactions to history study. Such a pupil, if tested by a scale which is exacting in its requirements as to reading and composition, will not be able to show his proficiency in history. The completion type of test seems to meet these conditions best.

Moreover, a history test should be so arranged and the directions and exercises should be so explicit that equivocal or ambiguous answers are not possible. No answer should call into exercise the judgment of the grader or scorer of the test. Such tests are also more easily and accurately graded with less expenditure of time and effort on the part of the teacher. It is even desirable that the exercises and scoring devices be such that pupils may score their own papers.

A consideration of the difficulties and possibilities of testing

in the field of history seems to indicate the following requirements of a good test:

1. It must cover a wide range of facts, those facts being ones which have been presented with emphasis. (Names, dates, places, periods, events, principles, terms, problems, causes, and results, etc.)

2. The test exercises must demand the use of a variety of mental functions and processes. (Memory, association, judgment, reasoning, organization, etc.)

3. It should in so far as possible be independent of attainment in other subjects; it should require little reading, composition, or spelling ability.

4. It should require unequivocal answers.

5. It should be easily given and easily scored.

One such test[3] which meets the above requirements has been devised and standardized by the writer. There are ten exercises each including from two to ten elements arranged in the order of their difficulty. The exercises vary in the types of responses called for and very little writing is required. A tabulation of the exercises follows:

Exercise I	Men-Achievements. Five completion exercises.
Exercise II	Terms-Examples. Ten exercises in understanding of terms.
Exercise III	Places-Events. Five places; to name historic events.
Exercise IV	Problems-Solution. Two situations; state type of action called for.
Exercise V	Periods-Men. Five periods; associate men with period.
Exercise VI	Events-Years. Five events; state year of occurrence.
Exercise VII	Dates–Events. Five dates; give significant events.
Exercise VIII	Principle–Proof. Four general statements; give example as proof.
Example IX	Movement–Significance. Four movements; state importance of each.
Example X	Causes–Results. Five causes; state results.

[3] Rugg, E. U. "Character and value of standard tests in history," *School Review*, 27:757-71, December, 1919.

Two points are given for each correct underlining, checking, or other form of answer and one point for partly correct answers. A key for scoring is also furnished. Thirty-five minutes is the time required to give the test, and pupils may score their own papers under the direction of the teacher.

The original test exercises were given to more than twelve hundred seventh- and eighth-grade pupils and to approximately six hundred high-school, normal-school, and university students. Returns were received from twenty-three schools in nine cities as follows: Detroit, Kansas City, Missouri, New York, Minneapolis, St. Paul, West Chester, Pennsylvania, Aberdeen, South Dakota, and Milbank, South Dakota. On the basis of the answers received the difficulty of each exercise was determined and tentative standards evolved. These standards need to be revised since they are based upon the percents of correct answers in the original one hundred exercises and not upon the fifty elements included in the test in its final form. They are, however, offered here as approximately correct. They represent median scores for classes at or near the end of the school year.

Grade VII	56
Grade VIII	86
Grades IX, X, and XI	72
High-school seniors	90
Normal-school students	96
University students	90

It will be noted that the median for the ninth, tenth, and eleventh grades combined is somewhat lower than that of the eighth grade. The medians are approximately the same for each of the first three high-school grades. No American history was taught in these high schools until the senior year. The seniors were completing their year or half year of study of American history. The normal-school students were also taking courses in American history. The university students had not studied American history since leaving the high school.

The writer does not contend that the test here described is a measure of all the results of instruction in history. Other information tests as well as "problem" tests and tests of appreciation should be devised and standardized and thus made available to teachers for measuring additional outcomes of their teaching of history.

Editorials

OUR CONTRIBUTORS

Our proximity to the grindstone during the past twelve months has been both prolonged and—touching, so to speak. As this number of the year goes to press we therefore find that we have been too preoccupied to realize our right to celebrate. Now, however, that our eyes are opened to the opportunity, our feelings are to say the least "mingled." We should like, however, to detach from them a sentiment of gratitude to our contributors, because it is they alone who have made it possible for us to have what we are pleased to call our "policy," or indeed to have a magazine at all.

Our contributors have been sympathetic. They have caught the idea; and they have sent us the sort of copy we wanted. Moreover, the material which they have submitted has steadily improved both in amount and quality. And it may confidently be expected to continue to do so. Articles in hand and promises received justify us in playing the rôle of prophet to this extent.

Indeed, in order to bring out additional valuable material, we found it advisable to enlarge our last two numbers to ninety-six pages. This was accomplished through the generosity of the publishers who, at a time when every square inch of paper had to be put to its best use, ignored the additional expense.

Our contributors have also been generous. They have overlooked a number of *gaucheries* on our part, doubtless—and we hope rightly—attributing them to growing pains. They have endured delays in handling manuscript—delays due, as they may have charitably supposed, to an editorial office not at first well organized for its work.

They have cooperated; they have magnificently acquiesced in our editorial supervision. We have tried to make it sympathetic, and they have been good enough to say we have succeeded. We have, however, been fussy. We have verified references, compared quotations with originals, reconciled text with tables, and checked up computations. We have accordingly made or

suggested to authors many alterations. In no case, however, has an author failed to accept a proposed alteration without giving a better reason for his version than we had given for ours. We have not yet encountered the author who believes that his self-respect is involved in maintaining his manuscript unaltered. (Perhaps there is only one thing to do with an article from such an author—namely, to reject it without reading it.) All the authors with whom we have come in contact have apparently looked upon the publication of their papers as joint enterprises in which both they and the editors were uniting to secure the best results.

We have therefore come down to the end of the year with an unbroken experience of pleasant relationships with our contributors. We cherish their sympathy; we appreciate their generosity; and we are grateful for their cooperation.

B. R. B.

THE NEXT STEP AFTER TESTING

Research workers have devised instruments for measuring attainment in school subjects; and teachers have used them by the million. They have administered tests, graded papers, and tabulated scores. In each school while this sort of work has been going on, hopes have been high. Sometimes, to be sure, it has not been the teachers who have hoped; but it is to be assumed that some one, or some group, has looked forward with faith to the time when conditions would be revealed.

Nor, as a matter of literal fact, have the tests in general failed to reveal the conditions. There is little disposition anywhere—as far as one may judge from printed reports—to gainsay the essential truth of the verdict which the tests may have pronounced. We are not blind to their shortcomings as measuring instruments, nor do we believe them to be possessed of the accuracy of either the yardstick or the spring balance. Yet we are confident that thoughtful school people do not expect such a degree of accuracy. Users of test material apparently believe that the results they have secured are essentially true.

But they expect something more—something, indeed, which no measuring instrument, as such, can be expected to afford. They ask, "What shall we do with these results after we have secured them?" This is distinctly another problem. It has

nothing to do with the tests as measuring devices—no more than the procedure of a physician has to do with his clinical thermometer.

We have heard it said that the test movement will fail unless it can improve instruction. The idea is good, but the expression is bad. So far as it is taken to mean that tests should improve instruction, it reveals a lot of muddy thinking. Tests were not devised to improve instruction but to measure its product. They should be judged by their validity and reliability as instruments for such a purpose. If by means of them we can make truer statements regarding educational outcomes than we can without them, if through their use we can adopt a common language and make our statements not only true but also indisputable, their usefulness is established. Whether they improve instruction then becomes a question of whether with more accurate knowledge at hand more effective teaching may be expected to take place.

Such a question answers itself. All the experience of the race is to the effect that the more we know about processes the better we can control them. If instruction has not been improved as the result of knowledge gained from testing, it is not the fault of the tests, for they have provided the facts and to provide facts is their sole office. It is not the fault of the test-makers who in virtue of being test-makers have rightly contented themselves with perfecting their instruments.

It is the fault of those whose business it is to utilize the knowledge. It is primarily the fault of the school people. It is true that research workers—not as test makers but as students of education—have not supplied either the light or the leading which might have been expected. But many of them have rightly felt that the best methods would have to be worked out in the day-by-day contact with class-room conditions. There is a certain almost insuperable difficulty in making from an office or professorial chair recommendations as to instructional procedure. Such recommendations when made from such sources are apt to betray their origin by their impracticability.

Under these circumstances it is the school people who must be heard from. The "next step" after testing is theirs. The question is not one of theory but of practice, and it can best be solved by those who are engaged in practice.

B. R. B.

Reviews and Abstracts

E. H. CAMERON, *Editor*

PACKER, P. C., Anderson, H. W., and Brueckner, L. J. *Survey of public school building requirements in Cleveland Heights, Ohio.* Cleveland, Ohio: Board of Education, 1920. 39 pp.

School building surveys have pointed out very clearly that the school plants of many of our American cities are developed without any definite plan or program. Such surveys have indicated that boards of education have very infrequently thought of a development of their school plants over a period of years. In many instances, school buildings have been erected with present needs only in mind. School building sites have been purchased only as the need for the immediate location of a building arose, and mere judgment rather than scientific analysis of the entire problem has been only too frequently the basis for extensions to the construction program.

In many cities parents have brought pressure to bear for the erection of buildings within their districts, regardless of the future maintenance problem and irrespective of future developments. As long as school building costs were relatively low, such haphazard selection of school building sites and the construction of buildings of varying sizes were not clearly censured by the taxpayers. With the addition of various kinds of special room facilities to school buildings, the establishment of higher standards of construction, the rapid increase in the cost of construction as well as the increase in cost of maintenance, the public has, in an increasing degree, been demanding of boards of education the development of scientifically prepared programs for the school building plant.

A second impetus has been given such scientific analyses due to the participation of large private organizations in the development of similar programs for their own physical plants. Telephone companies, chain-store corporations, municipal water companies, and the like, have all found it necessary to make extensive investigations into population growth and population trends before expending large sums in the extension of their plants.

This report on the public school requirements of Cleveland Heights is one of the relatively few scientific analyses that have been made by the smaller cities in the United States in order to show what the school building problem will be over a period of years.

The authors have based their program upon population studies made for the extension of subways and other service systems. They have set up standards for sites, the distances to be travelled by children in the various types of schools and have suggested the building program for immediate adoption, as well as the possible program which the board must bear in mind over a period of the next twenty years.

The report shows the building needs for elementary schools, and for junior and senior high schools. It must be clear to the taxpayer who reads this report that all

partiality in the selection of sites has been eliminated and that equality of service for all sections of the community has been the underlying motive.

The Cleveland Heights study should be duplicated for all cities in the United States. The insistence by school superintendents on the development of such a program for their cities will free them from much political annoyance. The adoption of a building program assures freedom from strenuous building campaigns every time that a new school building is needed, and also assures the selection of sites which will enable school principals to develop a program of play in conjunction with the other school work. The unbusiness-like method of selecting school sites only as the need occurs has resulted in reducing our playground areas to a most unfortunate minimum in most of our cities. It also has resulted in excessive prices for sites because of increased property values.

As more frequent studies are made in this field, the technic of the work will be gradually advanced. The authors, in this instance, have made little contribution to the technic which has previously been developed, but on the other hand, have shown a very clear appreciation of their problem and have utilized the best elements that are to be found in previous school building studies.

N. L. ENGELHARDT

BETELLE, JAMES O. *A checking schedule for projected school buildings.* Milwaukee: Bruce Publishing Company, 1919. 32 pp.

The author has drawn the subject-matter for this pamphlet from a wide experience in the planning and construction of school buildings. He, like other architects, has been aware of the confusion and complexity of the problem which confronts the architect in his contacts with the educational authorities of school systems. Most architects have learned that school boards and school superintendents have not clearly outlined their school building problem before the architect is called into consultation. As a result, architects are required to formulate floor plans of school buildings with only a partial knowledge of what the school authorities desire or consider essential. Thus, architects are frequently required to make floor plans over and over again in order to include those rooms or accommodations, the need for which appears only through the review of floor plans. Mr. Betelle has endeavored to eliminate unnecessary duplication of floor plans by presenting a checking list which will enable school boards and school superintendents to determine at the very outset what types of facilities they desire to have included within the school building which is being planned.

An architect who utilizes part I of this pamphlet in his first contacts with the board of education and the superintendent of schools will succeed in establishing better relationships with the school authorities than would otherwise result. He will also save himself and his clients money, time, and unnecessary worry.

Part II consists of a checking list for the specifications which accompany the final working drawings made by an architect. Mr. Betelle has included here a list of definite items on which a superintendent of schools may check the specifications of the architect. The author has recognized that because of the bulk of data included within specifications, neither the responsible member of the school board nor the superintendent finds it possible to study the specifications with care.

The items of this checking list appear to be those of greatest importance and include many elements which may affect the educational program. This pamphlet will serve a very definite need. If used by superintendents of schools, it will enable them to secure better school buildings than they otherwise would get. In most of our communities, school buildings are only built at intervals of two, three or more years. The superintendent who is confronted with a school building problem needs the assistance of other devices which will enable him to secure for his community a school building which best meets its needs. Mr. Betelle has made a contribution in this field which is a splendid beginning. An extension of his idea to the point where specifications and working drawings may be completely checked by the superintendents of schools now becomes most desirable.

N. L. ENGELHARDT

ALEXANDER, CARTER. *School statistics and publicity.* (Beverley Educational Series.) New York: Silver, Burdett and Company, 1919. 332 pp.

The purpose of this book is to help "the superintendent of schools, active or in training, to arouse his public and to secure adequate support for his school." The help is to be given by presenting the simpler effective working tools of the research student in education, and of the educational publicist.

The contents of the book may be grouped conveniently under four main heads: (1) the need for better school statistics; (2) the collection of school data; (3) educational statistics; and (4) the effective presentation of school data to the public.

Dr. Alexander's style is popular and pleasing. The book is one that will appeal especially to the busy superintendent who must be tempted by sweets as well as meats in the reading that he does. The technical phases of educational statistics that are presented are reduced to the simplest terms and are made quite clear by illustrations drawn from the superintendent's problems. For superintendents in training the book would serve well as a basic text in an introductory course.

The outstanding contribution of the book is the discussion of the problem of effective presentation of school data to the public. Not only are the chapters dealing with this problem of interest and of value to the superintendent in training but they will appeal very strongly to the experienced superintendent whether he has had training in educational statistics or not. This part of the book consists of an excellent collection of graphs, pictures, illustrative phrases, etc., that have been used by the leading educational publicists in this country in surveys, books, and magazine articles. The hundred pages treating this problem form an excellent reference book on graphic presentation of school statistics. For these pages alone every superintendent should own a copy of the book.

The book is not well proportioned. The presentation of school statistics to the public is the core of the book and its excuse for appearance. Only one hundred out of three hundred thirty-two pages are devoted to this treatment. Too much of the book is given over to the problems of why we need better school statistics and how to collect data—one hundred pages in all. One hundred pages are given to an elementary treatment of educational statistics. This part of the book would have been more usable as a text if the author had given many more problems for the student to work out.

The reviewer used the book as a text during the past summer and found that the students enjoyed and readily grasped the discussions. The preceding criticisms are an outgrowth of this experience. With supplementation and expansion of the chapters

on educational statistics and on presentation of school statistics to the public the book makes a very usable text of college grade for an undergraduate introductory course.

A. G. CAPPS

ELSON, J. C. and TRILLING, BLANCHE M. *Social games and group dances.* Philadelphia: J. B. Lippincott Co., 1919.

The development of community recreation has created a need for concrete material, a need especially felt by leaders who have not had special training in play. Dr. J. C. Elson and Miss Blanche Trilling have made a splendid contribution to this field in their recent book *Social Games and Group Dances.*

The book is divided into three parts: (1) social mixers and acquaintance games, active tag games, parlor games and "stunts"; (2) group dances; (3) singing games and dances for children.

Games and dances are described which are suitable for children and adults, for small and large groups, and for various social occasions. The acquaintance games are especially well selected both as to variety and originality. They are very "playable." The "stunts" are clever and well adapted for small groups. The descriptions of the dances are accompanied by the music and illustrations of the different figures, and can be easily learned.

Undoubtedly the best recommendation that can be given is the fact that these different forms of recreation presented have been successfully "tried out" by the authors who have had wide experience in the teaching of recreation. Any leader in recreational work may do well to have this book as a guide. It furnished material for the kind of recreation which is most valuable from the standpoint of wholesomeness and real fun.

LOUISE FREER

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Walter S. Monroe, University of Illinois, Urbana, Illinois.

South Dakota Educational Research Association

About two hundred teachers and superintendents attended the first meeting of the South Dakota Educational Research Association which was held at Aberdeen, South Dakota, June 22–23, 1920. The membership in this association is announced as follows:

City Superintendents	20
City Teachers	75
Rural Teachers	150
County Superintendents	5

This membership is interesting because of the relatively large number of rural teachers.

A Study of Rural Schools

A committee appointed by the Department of Rural Education of the National Educational Association is organizing a study of the following questions: "In which of the two outstanding types of rural schools, the consolidated and the one-teacher, are the results of instruction superior?" The study is to be made by giving the following educational tests: Monroe's Standardized Silent Reading Test, Woody-McCall Fundamentals in Arithmetic, Trabue Language Scales B or C, Iowa Spelling Scale, and Ayres Handwriting Scale.

A state director has been appointed by each state to cooperate with the committee in this work. Any county superintendent who is interested should communicate with the chairman of the committee, J. M. Foote, State Department of Education, Baton Rouge, Louisiana.

Supervision of Testing in Ohio

Director R. L. Morton of the Extension Department of Ohio University, Athens, Ohio, is initiating a testing program in southeastern Ohio. The tests are to be given on the following dates: first testing, Tuesday, October 12 to Friday, October 15, inclusive; second testing, Tuesday, January 18 to Friday, January 21, inclusive; third testing, Tuesday, May 3 to Friday, May 6, inclusive.

The tests recommended for use are Courtis Standard Research Tests in Arithmetic, Series B, Monroe's Timed Sentence Spelling Test, Monroe's Standardized Silent Reading Tests, and Ayres' Handwriting Scale.

Director Morton has issued detailed directions covering the ordering of the test material and the use of it in the schools. In addition he places at the disposal of the superintendents and teachers his department. Where possible he plans to visit the schools where assistance is desired and demonstrate the uses of tests in classes of children.

Superintendents' Conference at Illinois

The first annual conference of the superintendents of city schools of Illinois was held at the University of Illinois, November 17 and 18, 1920. This conference was organized under the supervision of the Bureau of Educational Research of the College of Education.

The library division of the Bureau of Educational Research exhibited bulletins, recent educational and intelligence tests, textbooks, and magazines. This exhibit was maintained not only for the benefit of the Superintendents' Conference but also for the High School Conference held during the same week.

As the following program indicates, the conference was largely devoted to the consideration of educational research.

Wednesday, November 17, 9:30 A. M.
Professor B. R. Buckingham of the
University of Illinois, Presiding

Music.....Urbana Public Schools
Address of Welcome.....President David Kinley
Address: "Scientific Methods of Improving Reading".....Professor C. H. Judd
Director of School of Education, University of Chicago
Address: "Individualization of Instruction".....Professor E. H. Cameron
Professor of Education, University of Illinois

Address: "A Recent City-wide Project in the Schools of Decatur". Mr. J. O. Engleman
Superintendent of Schools, Decatur
Luncheon 12:15 at Trinity M. E. Church

Wednesday, November 17, 2:00 P.M.
Superintendent T. W. Callahan of
Galesburg, Presiding
Address: "A Fair Deal for City Schools under Unit Control".....Mr. G. P. Randle
Superintendent of Schools, Danville
Round Table: "Scientific Methods of Improving Reading".....
.....Led by Professor C. H. Judd
Address: "Physical Education Legislation".....Mr. E. D. Caulkins
Manager, National Physical Education Service, Washington, D. C.
Get-Together-Dinner, 6:15 P. M. at Trinity M. E. Church

Wednesday, November 17, 8:00 P. M.
Superintendent D. Walter Potts of
East St. Louis, Presiding
Music.....School of Music, University of Illinois
Address: "Cooperative Research".....Professor B. R. Buckingham
Director of Bureau of Educational Research, University of Illinois
Address: "Projects Being Carried on by the Bureau in Chicago"..Mr. P. R. Stevenson
Bureau of Educational Research, University of Illinois
Address: "Reorganization of the Upper Grades of the Elementary School".....
.....Professor C. H. Judd..

Thursday, November 18, 9:30 A. M.
Superintendent E. E. Lewis of
Rockford, Presiding
Music.....Urbana Public Schools
Address: "Intelligence and School Progress".....Professor M. E. Haggerty
Dean of School of Education, University of Minnesota
Address: "Individual Versus Group Standardization"....Professor Walter S. Monroe
Assistant Director, Bureau of Educational Research, University of Illinois
Round Table: "Intelligence and School Progress".Led by Professor M. E. Haggerty
Luncheon and Business Meeting, 12:15 P. M. at
Trinity M. E. Church

Thursday, November 18, 2:15 P. M.
Superintendent E. C. Fisher of
Rock Island, Presiding
Address: "The Essential Criteria of a Satisfactory Test"...Professor M. E. Haggerty
Address: "The Use of Educational Tests as a Basis of Promotions, Groupings and Stimulus to Pupils and Parents."".....Mr. I. M. Allen
Superintendent of Schools, Springfield
Address: "Measurement as an Aid in Supervision".....Dean C. E. Chadsey
College of Education, University of Illinois

The following communication dated October 1, 1920, has been received from J. Carleton Bell, Secretary-Treasurer of the New York Society for the Experimental Study of Education. Because this announcement is of general interest it is printed herewith.

Society for the Experimental Study of Education

"Are you interested in improving the efficiency of teaching? Do you want to know what mental equipment your pupils bring with them to your classes? Could you not help them more if you knew their capacity for learning in various fields?

"The New York Society for the Experimental Study of Education is an organization designed to afford a clearing house of ideas for those who are interested in such questions. It meets at the Washington Irving High School on the second Friday evening of each month from October to May. The annual dues are $2.00. For the current school year special emphasis will be laid on the problems of the high school. In spite of the obvious need for the most careful, scientific studies of the mental capacities of high school pupils, their learning processes, and their methods of attack upon the subject matter of their courses, little experimental work has been done in this field. It is hoped that the high school teachers of New York City will appreciate the importance of such studies, and that each one will desire to take part in some investigation. Your attention is called to the problems suggested in the enclosed Bulletin.

"Do you not think that this movement is worthy of your support? Even though you may not be able to attend the meetings of the Society and participate actively in its work, do you not want to identify yourself with it, receive its Bulletins, and help in the furtherance of its aims?

"May I not present your name for membership at the next meeting? Please let me hear from you in regard to this matter."

Improvement in School Subjects Throughout the School Year

Dr. Henmon's article "Improvement in School Subjects Throughout the School Year" in the February number of the *Journal of Educational Research*, deals with facts of so much importance to the subject of educational research that his conclusions merit great consideration. I venture to suggest that some of his conclusions are not warranted by his own data. In dealing with the subject of individual differences in amount and rate of improvement, Dr. Henmon says that he believes that his data unmistakably point to the conclusion that individual differences tend to increase rather than to decrease with equal amounts of practice. Referring to his own data, he says that "an inspection of the tables shows that by any rational method of interpretation, the scriptural law 'to him that hath shall be given' holds. Those with the highest initial scores have the highest final scores and gain most both absolutely and relatively." The tables undoubtedly do show this, but I question the method by which they were constructed. This method was to arrange each grade tested into groups (quartile groups for grade IV and tertiles for grades V and VI), on the basis of the final, or May, scores of the pupils of that grade. The scores of the individual pupils were then traced back through the preceding months, the grouping made on the basis of the May scores being retained. According to this method it is found, for all three grades, that the groups which made the best average scores in May made also the best average scores in September. It is also found, and Dr. Henmon believes this to be very significant, that the amount of improvement of the superior groups is greater than is the amount of the improvement of the inferior groups.

It seemed to me, upon reading Dr. Henmon's article, that at least as good a way to group the pupils, would be according to their initial, rather than their final scores, and I wondered if such grouping would yield the same results as he found. To test this out, I have worked over his data, making my grouping on the basis of the September[1] rather than the May scores. My Group I consists of individuals who made up the top quartile (or tertile, for grades v and vi) of the *initial* scores. The average deviation of the scores from the class average has been calculated for each group in the same manner as in the original article. These deviations are given in the accompanying tables. A glance at these tables shows that when the pupils of Dr. Henmon's study are grouped according to their initial score, differences between pupils do not tend to increase with practice. In fact, in the case of all three grades, the *absolute* gain of the group with the lowest initial score was greater than that of the group with the highest initial score. The *relative* gain, of course, is even more in favor of the low group. When the scores are expressed in terms of deviations from the class averages, it is seen that the extreme groups apparently become more, rather than less, alike, from the beginning to the end of the school year.

It is not my intention to claim that these data actually point to a decrease in difference with practice. As Dr. Henmon says, the problem of measuring individual differences in improvement is very complex. Differences in units of measurement, as well as other complications, are quite probably great enough to swamp the small tendency I have here shown to exist (when pupils are grouped according to *initial* ability) towards an increase in "likeness." It is, however, claimed that Dr. Henmon's conclusions are unwarranted, unless he is able to show that his method of grouping is

TABLE I. GROUP DEVIATIONS FROM CLASS AVERAGES. PUPILS GROUPED ON BASIS OF SEPTEMBER SCORE. GRADE IV

	Groups			
	I	II	III	IV
September deviations.....	6.0	0.4	−1.8	−4.6
May deviations.....	4.5	−3.5	−0.7	−3.1

TABLE II. GROUP DEVIATIONS FROM CLASS AVERAGES. PUPILS GROUPED ON BASIS OF SEPTEMBER SCORES. GRADE V

	Groups		
	I	II	III
September deviations.....	7.3	−1.0	−5.3
May deviations.....	3.8	+1.8	−5.7

[1] The October rather than the September score is used for grade vi, because Dr. Henmon, in his article, omits the September scores of three pupils in this grade. A pupil's score is the sum of the number of examples right on the Courtis Standard Research Tests in Arithmetic, Series B.

TABLE III. GROUP DEVIATIONS FROM CLASS AVERAGES. PUPILS GROUPED ON BASIS OF SEPTEMBER SCORES. GRADE VI

	Groups		
	I	II	III
September deviations	9.8	−1.0	−8.8
May deviations	5.7	6.1	−11.7

a better one than the one which I adopted, for solving this problem. For practical purposes of prognosis it is surely more useful to evaluate pupils according to initial, rather than according to final achievement.

KATHERINE MURDOCK
Instructor in Psychology, Extension Department, Columbia University

A Bright Boy

Harold M. F. was born February 24, 1916. He was tested by the writer with the Stanford Revision of the Binet-Simon Tests, April 15, 1920. His age that day was 4–2. His mental age was found to be 7–0. His intelligence quotient, after the manner of Terman, was 168.

The test was made in the office of the superintendent of schools of the small city where Harold lives. The superintendent, the high-school principal, and both of the subject's parents were present and observed the test. There was no indication that the subject was affected by the presence of these four people. There was, however, one distracting influence. On the wall of the office hung a large clock which controlled the ringing of class-period bells in the building. The occasional peculiar noises made by the mechanism of this clock attracted and held his attention. Only after the examiner had given a brief explanation of the clock and its function was he willing to give heed to the tests.

We had heard of the unusual performances of Harold on a previous visit to this city and had no difficulty in arranging through the superintendent of schools for the test. His parents were very much interested in the tests and their outcome and gladly supplied all the information desired concerning the boy's activities and interests.

The examination began with the four-year level. He readily passed all of these tests. He also passed all but the aesthetic comparison test in the five-year group, and all but the mutilated pictures test in the six-year group. He missed the bow knot and diamond tests in the seven-year group, and the comprehension, definitions, and vocabulary tests of the eight-year group; and he passed one, the date test, in the nine-year group. In this last test he gave correctly the day of the week, the month, the day of the month and the year. His parents say that he has at home a perpetual calendar which he adjusts each morning.

Harold has a sectional map of the United States which interests him very much. The map is cut along the state lines, there being but one state to a section except that in the eastern part of the Union two or three small states constitute a section. He places

the sections by first reading the names of the states and then laying each section in approximately the correct position even though that position be, for a moment, isolated. "Take the gentleman on a trip to Texas," suggested his father. Harold replied, "Kentucky, Tennessee, Mississippi, Louisiana, and Houston or Galveston, Texas." (Harold lives in Ohio.) Similar trips to other parts of the country were suggested and correctly "taken."

The boy's keen interest in the facts of place-geography is remarkable. He gave correctly the counties in which the following Ohio cities are located:

Dayton	Portsmouth	Athens
Akron	Zanesville	Marietta
New Lexington	Columbus	Youngstown
Chillicothe	Cleveland	Sandusky
Cincinnati	Newark	Toledo
Springfield	Mt. Sterling	Lisbon

Furthermore he spells the names of these cities and of the counties in which they are located. The counties in order are: Montgomery, Summit, Perry, Ross, Franklin, Cuyahoga, Licking, Madison, Athens, Hamilton, Clark, Scioto, Muskingum, Washington, Mahoning, Erie, Lucas, and Columbiana.

He can name the capitals of all the states and can spell each. The writer had heard of this accomplishment in advance of the test but tried him on a dozen or so, located in all parts of the country. No errors were made.

But Harold's knowledge of city names and locations extends beyond the capitals. With one exception he gave correctly the states in which the following cities are located:

Newark	Tacoma	Cincinnati	San Francisco
Baltimore	Dubuque	Cleveland	Detroit
Erie	Houston	Evansville	Pasadena
Charleston (S. C.)	Oklahoma City	Chicago	Galveston
Louisville	New Haven	Birmingham	Minneapolis
Chattanooga	Pittsburg	Kalamazoo	Milwaukee
Dayton	Savannah	Parkersburg	New Orleans
Akron	Buffalo	Portland (Ore.)	St. Louis
Marietta	Petersburg	Portland (Me.)	Los Angeles
Omaha	Portsmouth	Jacksonville	Colorado Springs

We have copied this list from a slip of paper on which he had previously written them, reproducing them as he prepared them, spelling included. The one failure was Louisville, Kentucky. He cannot write script but prints his words.

The writer has a list of 25 birds which his mother says he readily recognizes when their pictures are encountered. Recently he attended with his parents a meeting of a local club called "The Science Club." A club member read a paper on the subject of birds and used a projection lantern in picturing the common American birds. As the birds came on the screen Harold amused the assembly by loudly announcing their names. This incident is vouched for by the superintendent of schools.

Harold has recently become interested in the names of automobiles. The writer has a list of 23 common cars, arranged alphabetically, which he has evidently copied from an advertisement, which he is learning, including the spelling.

At the age of two years and nine months Harold could tell the time of day. At this time he was presented with a watch which he carried for a number of months, proudly yet seriously informing his associates of the time. As a test we asked him to tell us the

time by the clock which had attracted his attention. He replied, "Eight minutes after ten." This was within 30 seconds of being exactly correct. The precision of his statement is worthy of note. He spells the names of all the days of the week and months of the year and can correctly give the number of days in each month. He also reads the thermometer and has been known to comment on the danger of frost when the thermometer goes below 40.

This unusual boy can read and write numbers to a thousand. We handed him a copy of Terman's *The Measurement of Intelligence* and asked him to turn to page 141 and then to page 267. He readily did so. We have a paper containing a long series of three place numbers which he has written. All are legible and fairly neatly done. Furthermore he writes the Roman numerals to C, as attested by a list of them in our possession.

His interest in mechanical things is shown by the projects which he sets for himself with children's building blocks. He has a large supply of blocks of various sizes out of which he has constructed buildings in imitation of the various state capitols, pictures of which he has seen. At present he is engaged in constructing the rather elaborate front of a large hotel in imitation of the picture which he has on a postal card.

Harold has evidently taught himself to read. His mother, seemingly a very intelligent woman, is unable to tell how he learned. She relates that before he was two years of age his grandfather frequently amused him by reading to him stories from children's books. Soon thereafter his parents found him reading for himself. He read as well as the average pupil at the end of the first year in school from the *Story Hour Book, Book One*, one of his own books. We supplied him with a copy of the *Baldwin and Bender First Reader* and also the *Winston First Reader*. He read from each of these at sight as well as from his own *Story Hour Book*. He is in the habit of getting the evening paper when it is delivered and reading to his mother the weather forecast for the following day.

Harold had learned all the letters of the alphabet from alphabet blocks by the time he was 18 months of age. He could name all except W and L, which he was unable to pronounce. But when asked to locate these letters he showed no hesitation in identifying them.

The boy's father is a college graduate, has studied law, has been prominently connected with two banks and has been admitted to the bar. He was a cashier in one of the banks. There seem to be no members of the family, immediate or remote, who have attained eminent distinction. Harold is the only living child in the family, a girl, who seems to have been very intelligent, having died before his birth.

The above incidents have been related to substantiate the findings of the Binet test. Further examples of the unusual performances of this remarkably precocious child might be supplied. All the evidence which the examiner was able to gather suggests that here may be a genius of a rare type.

Of course Harold's unusual performances cannot be attributed to training. Four-year-old children cannot be trained to do what this boy does easily unless they are possessed of an exceptionally high order of intelligence. Furthermore, the parents of this child have tried, unsuccessfully, to discourage him in his studious habits.

Harold is a healthy, vigorous, physically normal boy in every way. With the single exception of Spanish influenza which he had during the recent second epidemic, he has no disease history. He is the kind of boy one would gladly and proudly call one's own.

In the words of Terman, "How impatiently one waits to see the fruit of such a budding genius!"

R. L. MORTON
Ohio University

No Fairy Tale: The Need for Standardized Tests[1]

What are the two questions about your work that you would most like to have answered? Unless you are different from most good teachers, these questions probably are (1) How well are my pupils doing in their work? (2) How well ought they to do? The answers to these two questions give all good teachers grave concern. This concern becomes even greater when it is known that deficiency in certain school work means certain failure. When, for example, you know that unless your pupils have a certain capability in silent reading they will surely fail to complete the elementary school, and when you know that lack of a certain higher minimum ability in silent reading means failure to complete high school, and that lack of a still higher minimum means failure to complete college, then your concern for the progress of your pupils becomes grave indeed.

Measuring the results of your teaching will help you to know how well your pupils are doing; and the use of standards will help you to judge how well they ought to do. Good teachers have always been anxious to measure the results of their teaching. They have tried to do this by the use of examinations. Their success has depended on the kind of examinations that they have used, on the accuracy with which they have marked the questions, and on the use they made of the papers after they are marked. Because you have not always done well in your use of examinations is not necessarily a just cause for blaming you. Physicians did not do very well with fever cases before they had the thermometer, and they did not do well with lung trouble before they had the stethoscope. Before the invention of the X-ray many a person was afflicted from causes that are now promptly located and remedied.

It is probable that your teaching does improve because of your use of examinations. Before the physicians had thermometers they tried as best they could to measure the amount of temperature of a patient by such means as passing the hand over the head, noting breathing, dryness of lips, etc. So you note the daily work of your pupils, grade their papers, and give them examinations. Just as the physician improves his work through such better instruments as the thermometer, so you may improve your work through the use of better examinations. These better examinations are called *standardized* tests and scales.

Just as the physician did not know how much more he might be helping his patient before he had the X-ray, so we as teachers often do not know how much more we might be helping our boys and girls. One point that we shall need to see more clearly, then, is the disadvantage of the teacher-made or principal-made, or even the superintendent-made examination.

The next lesson will take up this point. Before reading the next lesson, write out your best answers to the following.

QUESTIONS AND EXERCISES

1. List at least three other instruments for accurate measuring besides the stethoscope and thermometer.

[1] Being the introductory lesson to a first course in Educational Measurements, now offered through correspondence. The title is suggested by that of Professor Mead's article in the September, 1920, number of the *Journal of Educational Research*.

2. What units of measure does the farmer use? The merchant? How are these standardized?

3. State some advantages of the use of standardized instruments for measuring potatoes.

4. What are the disadvantages of what is called barter in the trade of primitive peoples?

5. Fill in the blanks in the following sentences so as to make them state a summary of this lesson.

All good teachers are greatly concerned to know how..... their pupils are.....and how much..... they ought to..... ..... The best way to know is to..... Until recently the best means of knowing has been the .... Standardized tests are merely improved..... Just as the instruments which the physician uses for examining his patients are being improved, so the instruments which the teacher uses ought to be.....

C. W. STONE
State College of Washington

A Layman's Attitude Toward Intelligence Tests

Within recent years the use of standardized tests in school subjects and the use of tests for determining the intelligence of school children have increased with an astonishing rapidity. The opposition of educators which was prominent in our educational periodicals and at educational meetings has practically disappeared. An educator who has not been converted to their use is an exception. In the midst of such "prosperity" it is, perhaps, wholesome to note what may, possibly, be the attitude of a considerable number of laymen. We quote below at some length an editorial published in the *Peoria Transcript* for September 16, 1920. Presumably, the writer of the editorial is a layman.

According to plans promulgated by Dr. Walter S. Monroe of the bureau of educational research, University of Illinois, tests to determine the mental capacity of approximately 200,000 public school children of Illinois will be made during the week of October 4.

These tests are sponsored and for the most part devised by Dr. Monroe. Children of the first and second grades, it is announced, will not be examined. Dr. Monroe is quoted as saying that county school superintendents of nearly half the counties of the state have consented to the tests, and that classification and standardization of the children as individuals rather than by groups will be sought. The examination will consist of the three distinct scales; a group intelligence scale, "Monroe's Standardized Silent Reading Tests" and arithmetic tests.

We have no doubt that interesting statistics could be collated [*collected*] through the examination for mental age of 200,000 school children, but we seriously doubt the authority of county superintendents of schools to consent to the Monroe tests or any other device for determining the individual mental age of school children.

The public school is a constitutional institution created for the sole purpose giving all children of the state "a good common school education." It is supported by public taxation. We know of no authority by which public school children may be subjects for psychological clinics by a department of the state university.

We are living in an age in which the individual is called upon to make many sacrifices for the public good. Psychology is a comparatively new science and its devotees are fired with a fanatical zeal to broaden its boundaries. We are not prepared to say that the Monroe tests should be barred from the public schools, but certain it is that parents are under no legal or other obligation to submit to them.

It should be made clear in every county whose consent is given by its superintendent of schools that compliance is voluntary and that any public school pupil may be excused from the tests upon individual or parental request.

One erroneous statement in the above editorial should be pointed out. The testing program to which he referred was not initiated by the Bureau of Educational Research of the University of Illinois, but was initiated by the county superintendents, and the Bureau of Educational Research has acted only n an advisory capacity.

Dr. L. P. Ayres Leaves the Russell Sage Foundation to Enter the Cleveland Trust Company

Dr. Leonard P. Ayres, a director of the Russell Sage Foundation of this city, has been elected Vice-President of the Cleveland Trust Company of Cleveland, Ohio, and will take up his new work at once. Since the establishment of the Russell Sage Foundation in 1907 Dr. Ayres has been the Director of its Department of Education, and also of its Department of Statistics. The work of the Foundation in these two fields has attracted wide attention. Dr. Ayres is the author of a number of books on educational, statistical, and scientific subjects.

During the war he organized and directed simultaneously the statistical offices of the Council of National Defense, the War Industries Board, the Priorities Committee, and the Allied Purchasing Commission. He then entered the army and became its Chief Statistical Officer, and organized the Statistics Branch of the General Staff. He was commissioned as a Colonel, and established the statistical services at General Pershing's headquarters in France, as well as those of the Supply Service of the Expeditionary Forces. After the armistice he was appointed Chief Statistical Officer of the American Commission to Negotiate Peace, and accompanied President Wilson on his mission to France. For these services he was awarded the Distinguished Service Medal.

Dr. Ayres is a graduate of Boston University and took his postgraduate work at Columbia University. His educational writings deal largely with administrative and financial problems. Some years ago he made an analysis of large numbers of letters, newspapers, and other sorts of English composition, to identify the 1000 commonest words in English writing. The results of this research largely revolutionized the teaching of spelling in the public schools. In cooperation with his associates in the Russell Sage Foundation Dr. Ayres has conducted a large number of surveys of city school systems. Among his recent publications, his comparative study of state school systems and his book entitled, "The War with Germany," have attracted wide attention. Dr. Ayres will take charge of the research and statistical work of the Cleveland Trust Company.

National Association of Directors of Educational Research

(E. J. ASHBAUGH, *Secretary and Editor*)

I recently attended a meeting of school men at which a gentleman who had evidently spent some time in the study of the measurement of mentality dwelt at some length on the individual differences in the mentality of children and the possibility of discovering these differences by the use of general intelligence tests. He portrayed vividly the wide ranges of capacity found in the ordinary class of school children and the consequent fact of difficulty in teaching them as a single group. His major emphasis, however, was upon the superintendent's responsibility for the discovery of the mental capacity of the individual children within his school system.

When the meeting was over, the superintendents and principals in discussing the address seemed to generally arrive at one conclusion, voiced by one of their number when he said, "Perhaps any of us could discover the mental capacities of our children, but the more difficult and vastly more important problem is, what shall we do for them after they have been discovered. On that point he gave us no help."

This experience has doubtless been duplicated by many of the readers of this column. It raised a question in my mind, however, from which I shall have no rest until I pass it on to you, especially to you who are members of our research association. One of the privileges bestowed upon the secretary in return for his annoyance in endeavoring (often without avail) to get you to report your activities, is that of being able at pleasure to put to you the questions which bother him.

A number of you have been using the group tests for measurement of intelligence. Some of you have been reclassifying pupils on the basis of these tests; placing some in slow classes, some in fast groups, some in unclassified groups, etc. But what modification of the course of study have you made? What evidence have you that the new material which you have assigned them is better fitted to their capacities than the old? What reason have you to think that they may now, working at their own highest level of achievement, succeed and therefore be progressing and happy?

But another question further. What have you done as an individual; what has the association as a whole done; what has anyone done to enable the superintendent or teacher in a small school system, a system of 200 to 500 children in all twelve grades, to care better for the children after having discovered their capacities? Does the problem challenge your careful thought? Or does it appear unimportant to you as you face your problems in your own field?

Will you write your reactions to these questions to your secretary for publication in these columns?

But two communications from members have reached my desk during the past sixty days. They are as follows:

Miss Henrietta V. Race, recently of the Louisville schools, has accepted the position of Director of Mental Measurement in the Kansas City Schools. Miss Race while in Louisville established a psychological laboratory in 1914. During the intervening six years, she made two surveys of the city schools in reading, arithmetic, and language, and did some work in penmanship.

She examined 4,000 boys and girls in the white high school in reasoning tests in history and the freshman classes of the white and colored schools in language, arithmetic, spelling, and composition. She also gave the Army test to these freshmen.

In 1916 she organized one of the first classes, if not the first class for gifted children, in the United States, the class being organized on the basis of mental tests. Two other similar classes were organized later.

She faces the problem at Kansas City of mental measurement for complete classification.

The other communication was from Charles D. Dawson, assistant superintendent of schools in charge of educational tests and measurements. He states that his bureau is engaged in the following work:

1. Measuring of locational geography work by means of Courtis' test.
2. Applying group psychological tests to grades in several schools.
3. Using the Binet-Simon mental tests in individual cases.

We are glad to welcome as new members of our association the following:

Mr. D. Edgar Rice, Consultant, Division of Tests and Standards, Camp Grant, Illinois.

Dr. Herschel T. Manuel, Colorado State Normal School, Gunnison, Colorado.

Mr. James L. McCrory, Assistant Superintendent in Charge of Administrative Research, Omaha, Nebraska.

Mr. Edward McMillan, Research Examiner, Central High School, Omaha, Nebraska.

JUL 1921

Standard Educational Tests

Realizing the great demand for standard educational tests, this house has built up a list of tests comprising achievement examinations, intelligence examinations, and prognosis examinations. As additional matter is available which meets new requirements, it will be added to this list. It is very obvious that public school educators realize the value of standard tests which are being made an essential part of the school equipment.

All tests listed below provide all material required for giving the tests, writing them, interpreting the results and scoring.

COURTIS STANDARD PRACTICE TESTS IN ARITHMETIC. This device is showing marked increase in popularity and has been part of the regular equipment of a number of school systems for several years. The drill cards are furnished in convenient cabinets—three sizes, depending upon the size of the class. In addition there are tissue paper pads on which the pupils work their drills and enter their scores; the manual gives complete directions for giving the tests. In addition to furnishing scientific and efficient drill, the device provides for research and supervisory testing—a feature introduced in the 1920 Revision. Devised for grades 4 to 8.

HAGGERTY INTELLIGENCE EXAMINATION. These tests have proved unusually popular and the demand is rapidly increasing. They are designed for grades 1 to 3 (Delta 1) and 4 to 9 (Delta 2). They are especially adaptable to the usual school routine since the tests can be given in less than 30 minutes.

HAGGERTY READING EXAMINATION. This test was formerly known as "Achievement Examination in Reading" of which Sigma 1 (grades 1 to 3) has been available for the past eight months. Sigma 3 (grades 5 or 6 and the early high school grades) is now available for the first time. Sigma 2 will be issued during the year. The unusual demand for Sigma 1 shows that there is great need for this kind of tests.

HANDSCHIN MODERN LANGUAGE TESTS. These are tests for the examination of achievement in French and Spanish in four-year high schools and colleges. There are two forms in each French and Spanish test.

HUDELSON ENGLISH COMPOSITION SCALE. This is a book giving a number of samples of composition actually written by school children with the values indicated. Compared with other scales, this one avoids the uneven steps between the various samples and thus prevents confusion to teachers untrained in the use of such devices. The book gives full direction for the use of the scale and also a score key. Designed for use in grades 4 to 12.

NATIONAL INTELLIGENCE TESTS. The National Research Council provided two scales and one form of each scale for the examination of pupils in grades 3 to 8 or 9. These tests were first available this fall and are already in use in the examination of about three hundred thousand pupils.

OTIS GENERAL INTELLIGENCE EXAMINATION. Especially designed for the examination of applicants for clerical and executive positions. It is also usable in vocational classes. The examination involves the principle of self-administration, thus enabling the examiner to test an individual or a group of individuals without any attention except for scoring.

OTIS GROUP INTELLIGENCE SCALE. This scale was extensively introduced at the very outset about a year ago when the Advanced Examination appeared. During the last six months, the Primary Examinations were issued. The growing demand and reordering of the tests is evidence of their merit, and also the fact that scientific testing is becoming a part of the regular school organization in many places. Both Primary and Advanced Examinations are furnished in two forms, A and B.

TERMAN GROUP TEST OF MENTAL ABILITY. A second edition of this test was necessary within three months after its appearance. It is distinctive from the other intelligence tests in that it is designed especially for high school and college students. The use of standard tests in high schools is growing at a very high rate.

WILKINS PROGNOSIS TEST IN MODERN LANGUAGES. The purpose of this test is to ascertain whether a student is qualified to take up the study of modern languages. Although it is designed to give these tests before the study of languages is taken up, it is possible and advisable to use them at any time during the year; thus the teacher is able to check the results of the one examined with his progress in the class.

The publishers invite correspondence with any one interested in tests and will be pleased to furnish further information and prices. In view of the heavy demand for testing materials of all kinds, it is advised that orders be placed as far in advance as possible.

WORLD BOOK COMPANY

Yonkers-on-Hudson
NEW YORK

2126 Prairie Avenue
CHICAGO

AMERICAN BOOK COMPANY

EDUCATIONAL PUBLISHERS

NEW YORK, 100 Washington Square
CINCINNATI, 300 Pike Street
CHICAGO, 330 East 22nd Street
BOSTON, 63 Summer Street
ATLANTA, 2 North Forsyth Street

Address all communications to the nearest office of the company. Descriptive circulars, catalogues, and price lists will be mailed on request to any address.

Your correspondence is solicited and will have prompt attention.

A MINIMUM PRICE FOR THE SERVICE

A monthly service which will aid every school to get the best results is what we can now offer in the *Journal of Educational Research*. It brings all the teachers into united study of school work as a continuous unit.

The school boards of every city can well make this Journal part of the equipment for each school building. The teachers will read it eagerly and make its articles the guide of common efforts to unify the progress of pupils through the school.

$4.00 a Year is the Subscription Price

Substantially bound in two volumes the numbers of the first year are available for school libraries. The price is $3.00 per volume or $6.00 for the year. Only a limited number are available, so order at once.

A good subscription combination is: The Journal of Educational Research (B. R. Buckingham, Editor-in-chief. Price $4.00) and School and Home Education (W. C. Bagley, Editor. Price $2.00).

For $5.00 we will send the two magazines for one year.

Send all subscriptions to

The Public School Publishing Company,
Bloomington, Illinois

"How Do We Use the Results After We Get Them?"

Superintendents sometimes ask that question after they have administered a series of standardized tests. If the answer is not readily forthcoming, they are tempted to blame the tests and to ask further, "What good are the tests, then?"

Consider the Gold Miner of '49:

In the days of which Bret Harte wrote the gold seeker took pick and shovel and pan, and sometimes a placer outfit, and found gold in the hills of the Sierra Nevada.

But, having found the gold he sought, did he ask, "How do I use it?" No, not he. Did he blame his pick, his shovel, his pan, his quicksilver for not telling him how to use his nuggets? No. He may have wasted his hardwon treasure in the questionable delights of the Barbary Coast; he may have founded a Leland Stanford University. At all events he didn't ask anyone how to use his gold.

The alert, resourceful teacher will always find a way to use the results derived from standardized tests whenever you provide those results for her. The tests are merely the picks, the shovels, the pans and the quicksilver,—the tools with which you find the golden facts you need to know about your schools. Your teachers will find a way to use these facts,—and they will not fritter this gold in any pedagogical Barbary Coasts, either.

General Survey Tests That Give the Facts Needed

The Illinois Examination—prepared by Walter S. Monroe and B. R. Buckingham in the Bureau of Educational Research of the University of Illinois—is a complete testing instrument for city or for rural schools.

This Illinois Examination was described in the September JOURNAL OF EDUCATIONAL RESEARCH. The cost is $40.00 per 1,000 for Examination I or for Examination II. I is used for grades 3, 4 and 5; II for grades 6, 7 and 8.

The Silent Reading Test used in the Illinois Examination is a new form of Monroe's Standardized Test. It is in this form an economically applied general survey. Printed separately this comes as a four-page test for grades 3, 4 and 5; and a four-page test for grades 6, 7 and 8. The cost is 80 cents a hundred copies for either test.

The Arithmetic Test used in the Illinois Examination is a new form devised by Dr. Walter S. Monroe as a

General Survey Scale in Arithmetic

Test I is for grades 3, 4 and 5.

Test II is for grades 6, 7 and 8.

The price of each is 80 cents per 100 copies.

This General Survey Scale will be found of particular value for a general survey test. It is quickly administered and easily scored.

Order all of these tests from

THE PUBLIC SCHOOL PUBLISHING COMPANY
BLOOMINGTON, ILLINOIS

New Books for the Professional Library

Bonser's THE ELEMENTARY SCHOOL CURRICULUM

suggests a natural and gradual transition from the unsatisfactory organization of work on a subject basis to the organization of a curriculum based on the normal activities of child life. $2.20

Keith and Bagley's THE NATION AND THE SCHOOLS

is a study of the application of the principle of Federal Aid to education in the United States. The Smith-Towner bill and the proposed Department of Education are considered in connection with such urgent present-day problems as illiteracy, the inadequacy of the rural school, teacher shortage, and health education. $2.25

Wilson and Hoke's HOW TO MEASURE

explains scales and standard tests in such a way that the teacher may make intelligent use of them for measuring the ability of the individual pupil and for testing the efficiency of methods of teaching. $2.20

The Best New and Recent Books for Teachers' Reading Circles

Clark: THE HIGH SCHOOL BOY AND HIS PROBLEMS..... $1.20
Clow: PRINCIPLES OF SOCIOLOGY WITH EDUCATIONAL APPLICATIONS..... 1.80
Curtis: RECREATION FOR TEACHERS..... 1.80
Davis: THE WORK OF THE TEACHER..... 1.60
Freeland: MODERN ELEMENTARY SCHOOL PRACTICE.... 1.80
Hollingworth: PSYCHOLOGY OF SUBNORMAL CHILDREN.. 1.80
McMurry: TEACHING BY PROJECTS..... 1.60
Norsworthy and Whitley: PSYCHOLOGY OF CHILDHOOD.... 1.80
Sechrist: EDUCATION AND THE GENERAL WELFARE..... 1.60

The Macmillan Company

New York Boston Chicago
Dallas Atlanta San Francisco